

1. Introduction: The Nature of Science and Physics
 1. [Introduction to Science and the Realm of Physics, Physical Quantities, and Units](#)
 2. [Physics: An Introduction](#)
 3. [Physical Quantities and Units](#)
 4. [Accuracy, Precision, and Significant Figures](#)
 5. [Approximation](#)
2. Kinematics
 1. [Introduction to One-Dimensional Kinematics](#)
 2. [Displacement](#)
 3. [Vectors, Scalars, and Coordinate Systems](#)
 4. [Time, Velocity, and Speed](#)
 5. [Acceleration](#)
 6. [Motion Equations for Constant Acceleration in One Dimension](#)
 7. [Problem-Solving Basics for One-Dimensional Kinematics](#)
 8. [Falling Objects](#)
 9. [Graphical Analysis of One-Dimensional Motion](#)
3. Two-Dimensional Kinematics
 1. [Introduction to Two-Dimensional Kinematics](#)
 2. [Kinematics in Two Dimensions: An Introduction](#)
 3. [Vector Addition and Subtraction: Graphical Methods](#)
 4. [Vector Addition and Subtraction: Analytical Methods](#)
 5. [Projectile Motion](#)
 6. [Addition of Velocities](#)
4. Dynamics: Force and Newton's Laws of Motion
 1. [Introduction to Dynamics: Newton's Laws of Motion](#)
 2. [Development of Force Concept](#)
 3. [Newton's First Law of Motion: Inertia](#)
 4. [Newton's Second Law of Motion: Concept of a System](#)
 5. [Newton's Third Law of Motion: Symmetry in Forces](#)

6. [Normal, Tension, and Other Examples of Forces](#)
7. [Problem-Solving Strategies](#)
8. [Further Applications of Newton's Laws of Motion](#)
9. [Extended Topic: The Four Basic Forces—An Introduction](#)
5. Uniform Circular Motion and Gravitation
 1. [Introduction to Uniform Circular Motion and Gravitation](#)
 2. [Rotation Angle and Angular Velocity](#)
 3. [Centripetal Acceleration](#)
 4. [Centripetal Force](#)
 5. [Fictitious Forces and Non-inertial Frames: The Coriolis Force](#)
 6. [Newton's Universal Law of Gravitation](#)
 7. [Satellites and Kepler's Laws: An Argument for Simplicity](#)
6. Work, Energy, and Energy Resources
 1. [Introduction to Work, Energy, and Energy Resources](#)
 2. [Work: The Scientific Definition](#)
 3. [Kinetic Energy and the Work-Energy Theorem](#)
 4. [Gravitational Potential Energy](#)
 5. [Conservative Forces and Potential Energy](#)
 6. [Nonconservative Forces](#)
 7. [Conservation of Energy](#)
 8. [Power](#)
 9. [Work, Energy, and Power in Humans](#)
 10. [World Energy Use](#)
7. Linear Momentum and Collisions
 1. [Introduction to Linear Momentum and Collisions](#)
 2. [Linear Momentum and Force](#)
 3. [Impulse](#)
 4. [Conservation of Momentum](#)
 5. [Elastic Collisions in One Dimension](#)
 6. [Inelastic Collisions in One Dimension](#)
 7. [Collisions of Point Masses in Two Dimensions](#)

8. [Introduction to Rocket Propulsion](#)
8. Temperature, Kinetic Theory, and the Gas Laws
 1. [Introduction to Temperature, Kinetic Theory, and the Gas Laws](#)
 2. [Temperature](#)
 3. [Thermal Expansion of Solids and Liquids](#)
 4. [The Ideal Gas Law](#)
 5. [Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature](#)
 6. [Phase Changes](#)
 7. [Humidity, Evaporation, and Boiling](#)
9. Heat and Heat Transfer Methods
 1. [Introduction to Heat and Heat Transfer Methods](#)
 2. [Heat](#)
 3. [Temperature Change and Heat Capacity](#)
 4. [Phase Change and Latent Heat](#)
 5. [Heat Transfer Methods](#)
 6. [Conduction](#)
 7. [Convection](#)
 8. [Radiation](#)
10. Thermodynamics
 1. [Introduction to Thermodynamics](#)
 2. [The First Law of Thermodynamics](#)
 3. [The First Law of Thermodynamics and Some Simple Processes](#)
 4. [Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency](#)
 5. [Carnot's Perfect Heat Engine: The Second Law of Thermodynamics Restated](#)
 6. [Applications of Thermodynamics: Heat Pumps and Refrigerators](#)

7. [Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy](#)
8. [Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation](#)

11. Oscillatory Motion and Waves

1. [Introduction to Oscillatory Motion and Waves](#)
2. [Hooke's Law: Stress and Strain Revisited](#)
3. [Period and Frequency in Oscillations](#)
4. [Simple Harmonic Motion: A Special Periodic Motion](#)
5. [The Simple Pendulum](#)
6. [Energy and the Simple Harmonic Oscillator](#)
7. [Uniform Circular Motion and Simple Harmonic Motion](#)
8. [Damped Harmonic Motion](#)
9. [Forced Oscillations and Resonance](#)
10. [Waves](#)
11. [Superposition and Interference](#)
12. [Energy in Waves: Intensity](#)

12. Physics of Hearing

1. [Introduction to the Physics of Hearing](#)
2. [Sound](#)
3. [Speed of Sound, Frequency, and Wavelength](#)
4. [Sound Intensity and Sound Level](#)
5. [Doppler Effect and Sonic Booms](#)
6. [Sound Interference and Resonance: Standing Waves in Air Columns](#)
7. [Hearing](#)
8. [Ultrasound](#)

13. Geometric Optics

1. [Introduction to Geometric Optics](#)
2. [The Ray Aspect of Light](#)
3. [The Law of Reflection](#)
4. [The Law of Refraction](#)

5. [Total Internal Reflection](#)
6. [Dispersion: The Rainbow and Prisms](#)
7. [Image Formation by Lenses](#)
8. [Image Formation by Mirrors](#)
14. Electric Charge and Electric Field
 1. [Introduction to Electric Charge and Electric Field](#)
 2. [Static Electricity and Charge: Conservation of Charge](#)
 3. [Conductors and Insulators](#)
 4. [Coulomb's Law](#)
 5. [Electric Field: Concept of a Field Revisited](#)
 6. [Electric Field Lines: Multiple Charges](#)
 7. [Applications of Electrostatics](#)
15. Electric Potential and Electric Field
 1. [Introduction to Electric Potential and Electric Energy](#)
 2. [Electric Potential Energy: Potential Difference](#)
 3. [Electric Potential in a Uniform Electric Field](#)
 4. [Electrical Potential Due to a Point Charge](#)
 5. [Equipotential Lines](#)
16. Electric Current, Resistance, and Ohm's Law
 1. [Introduction to Electric Current, Resistance, and Ohm's Law](#)
 2. [Current](#)
 3. [Ohm's Law: Resistance and Simple Circuits](#)
 4. [Resistance and Resistivity](#)
 5. [Electric Power and Energy](#)
 6. [Alternating Current versus Direct Current](#)
 7. [Electric Hazards and the Human Body](#)
17. Magnetism
 1. [Introduction to Magnetism](#)
 2. [Magnets](#)
 3. [Ferromagnets and Electromagnets](#)
 4. [Magnetic Fields and Magnetic Field Lines](#)

5. [Magnetic Field Strength: Force on a Moving Charge in a Magnetic Field](#)
6. [Force on a Moving Charge in a Magnetic Field: Examples and Applications](#)
7. [Magnetic Force on a Current-Carrying Conductor](#)
8. [Torque on a Current Loop: Motors and Meters](#)
9. [Magnetic Fields Produced by Currents: Ampere's Law](#)
10. [Magnetic Force between Two Parallel Conductors](#)
11. [More Applications of Magnetism](#)
18. Electromagnetic Induction, AC Circuits, and Electrical Technologies
 1. [Introduction to Electromagnetic Induction, AC Circuits and Electrical Technologies](#)
 2. [Induced Emf and Magnetic Flux](#)
 3. [Faraday's Law of Induction: Lenz's Law](#)
 4. [Motional Emf](#)
 5. [Eddy Currents and Magnetic Damping](#)
 6. [Electric Generators](#)
 7. [Back Emf](#)
 8. [Transformers](#)
 9. [Electrical Safety: Systems and Devices](#)
19. Electromagnetic Waves
 1. [Introduction to Electromagnetic Waves](#)
 2. [Maxwell's Equations: Electromagnetic Waves Predicted and Observed](#)
 3. [Production of Electromagnetic Waves](#)
 4. [The Electromagnetic Spectrum](#)
 5. [Energy in Electromagnetic Waves](#)
20. Wave Optics
 1. [Introduction to Wave Optics](#)
 2. [The Wave Aspect of Light: Interference](#)
 3. [Huygens's Principle: Diffraction](#)

4. [Young's Double Slit Experiment](#)
21. Introduction to Quantum Physics
 1. [Introduction to Quantum Physics](#)
 2. [Quantization of Energy](#)
 3. [The Photoelectric Effect](#)
 4. [Photon Energies and the Electromagnetic Spectrum](#)
 5. [Photon Momentum](#)
 6. [The Particle-Wave Duality](#)
 7. [The Wave Nature of Matter](#)
 8. [Probability: The Heisenberg Uncertainty Principle](#)
 9. [The Particle-Wave Duality Reviewed](#)
22. Atomic Physics
 1. [Introduction to Atomic Physics](#)
 2. [Discovery of the Atom](#)
 3. [Discovery of the Parts of the Atom: Electrons and Nuclei](#)
 4. [Bohr's Theory of the Hydrogen Atom](#)
 5. [X Rays: Atomic Origins and Applications](#)
 6. [Applications of Atomic Excitations and De-Excitations](#)
 7. [The Wave Nature of Matter Causes Quantization](#)
 8. [Patterns in Spectra Reveal More Quantization](#)
 9. [Quantum Numbers and Rules](#)
 10. [The Pauli Exclusion Principle](#)
23. Radioactivity and Nuclear Physics
 1. [Introduction to Radioactivity and Nuclear Physics](#)
 2. [Nuclear Radioactivity](#)
 3. [Radiation Detection and Detectors](#)
 4. [Substructure of the Nucleus](#)
 5. [Nuclear Decay and Conservation Laws](#)
 6. [Half-Life and Activity](#)
24. [Atomic Masses](#)
25. [Selected Radioactive Isotopes](#)
26. [Useful Information](#)

27. [Glossary of Key Symbols and Notation](#)

Introduction to Science and the Realm of Physics, Physical Quantities, and Units

class="introduction"

Galaxies are
as immense
as atoms are
small. Yet the
same laws of
physics
describe
both, and all
the rest of
nature—an
indication of
the
underlying
unity in the
universe. The
laws of
physics are
surprisingly
few in
number,
implying an
underlying
simplicity to
nature's
apparent
complexity.
(credit:
NASA, JPL-
Caltech, P.
Barmby,
Harvard-
Smithsonian
Center for

Astrophysics)



What is your first reaction when you hear the word “physics”? Did you imagine working through difficult equations or memorizing formulas that seem to have no real use in life outside the physics classroom? Many people come to the subject of physics with a bit of fear. But as you begin your exploration of this broad-ranging subject, you may soon come to realize that physics plays a much larger role in your life than you first thought, no matter your life goals or career choice.

For example, take a look at the image above. This image is of the Andromeda Galaxy, which contains billions of individual stars, huge clouds of gas, and dust. Two smaller galaxies are also visible as bright blue spots in the background. At a staggering 2.5 million light years from the Earth, this galaxy is the nearest one to our own galaxy (which is called the Milky Way). The stars and planets that make up Andromeda might seem to be the furthest thing from most people’s regular, everyday lives. But Andromeda is a great starting point to think about the forces that hold together the universe. The forces that cause Andromeda to act as it does are the same forces we contend with here on Earth, whether we are planning to send a rocket into space or simply raise the walls for a new home. The same gravity that causes the stars of Andromeda to rotate and revolve also causes water to flow over hydroelectric dams here on Earth. Tonight, take a moment to look up at the stars. The forces out there are the same as the ones here on Earth. Through a study of physics, you may gain a greater

understanding of the interconnectedness of everything we can see and know in this universe.

Think now about all of the technological devices that you use on a regular basis. Computers, smart phones, GPS systems, MP3 players, and satellite radio might come to mind. Next, think about the most exciting modern technologies that you have heard about in the news, such as trains that levitate above tracks, “invisibility cloaks” that bend light around them, and microscopic robots that fight cancer cells in our bodies. All of these groundbreaking advancements, commonplace or unbelievable, rely on the principles of physics. Aside from playing a significant role in technology, professionals such as engineers, pilots, physicians, physical therapists, electricians, and computer programmers apply physics concepts in their daily work. For example, a pilot must understand how wind forces affect a flight path and a physical therapist must understand how the muscles in the body experience forces as they move and bend. As you will learn in this text, physics principles are propelling new, exciting technologies, and these principles are applied in a wide range of careers.

In this text, you will begin to explore the history of the formal study of physics, beginning with natural philosophy and the ancient Greeks, and leading up through a review of Sir Isaac Newton and the laws of physics that bear his name. You will also be introduced to the standards scientists use when they study physical quantities and the interrelated system of measurements most of the scientific community uses to communicate in a single mathematical language. Finally, you will study the limits of our ability to be accurate and precise, and the reasons scientists go to painstaking lengths to be as clear as possible regarding their own limitations.

Physics: An Introduction

- Explain the difference between a principle and a law.
- Explain the difference between a model and a theory.



The flight formations of migratory birds such as Canada geese are governed by the laws of physics.
(credit: David Merrett)

The physical universe is enormously complex in its detail. Every day, each of us observes a great variety of objects and phenomena. Over the centuries, the curiosity of the human race has led us collectively to explore and catalog a tremendous wealth of information. From the flight of birds to the colors of flowers, from lightning to gravity, from quarks to clusters of galaxies, from the flow of time to the mystery of the creation of the universe, we have asked questions and assembled huge arrays of facts. In the face of all these details, we have discovered that a surprisingly small and unified set of physical laws can explain what we observe. As humans, we make generalizations and seek order. We have found that nature is remarkably cooperative—it exhibits the *underlying order and simplicity* we so value.

It is the underlying order of nature that makes science in general, and physics in particular, so enjoyable to study. For example, what do a bag of chips and a car battery have in common? Both contain energy that can be

converted to other forms. The law of conservation of energy (which says that energy can change form but is never lost) ties together such topics as food calories, batteries, heat, light, and watch springs. Understanding this law makes it easier to learn about the various forms energy takes and how they relate to one another. Apparently unrelated topics are connected through broadly applicable physical laws, permitting an understanding beyond just the memorization of lists of facts.

The unifying aspect of physical laws and the basic simplicity of nature form the underlying themes of this text. In learning to apply these laws, you will, of course, study the most important topics in physics. More importantly, you will gain analytical abilities that will enable you to apply these laws far beyond the scope of what can be included in a single book. These analytical skills will help you to excel academically, and they will also help you to think critically in any professional career you choose to pursue. This module discusses the realm of physics (to define what physics is), some applications of physics (to illustrate its relevance to other disciplines), and more precisely what constitutes a physical law (to illuminate the importance of experimentation to theory).

Science and the Realm of Physics

Science consists of the theories and laws that are the general truths of nature as well as the body of knowledge they encompass. Scientists are continually trying to expand this body of knowledge and to perfect the expression of the laws that describe it. **Physics** is concerned with describing the interactions of energy, matter, space, and time, and it is especially interested in what fundamental mechanisms underlie every phenomenon. The concern for describing the basic phenomena in nature essentially defines the *realm of physics*.

Physics aims to describe the function of everything around us, from the movement of tiny charged particles to the motion of people, cars, and spaceships. In fact, almost everything around you can be described quite accurately by the laws of physics. Consider a smart phone ([\[link\]](#)). Physics describes how electricity interacts with the various circuits inside the device. This knowledge helps engineers select the appropriate materials and

circuit layout when building the smart phone. Next, consider a GPS system. Physics describes the relationship between the speed of an object, the distance over which it travels, and the time it takes to travel that distance. When you use a GPS device in a vehicle, it utilizes these physics equations to determine the travel time from one location to another.



The Apple
“iPhone” is a
common
smart phone
with a GPS
function.

Physics
describes the
way that
electricity
flows through
the circuits of
this device.
Engineers use
their
knowledge of
physics to
construct an

iPhone with features that consumers will enjoy. One specific feature of an iPhone is the GPS function. GPS uses physics equations to determine the driving time between two locations on a map. (credit: @gletham GIS, Social, Mobile Tech Images)

Applications of Physics

You need not be a scientist to use physics. On the contrary, knowledge of physics is useful in everyday situations as well as in nonscientific professions. It can help you understand how microwave ovens work, why metals should not be put into them, and why they might affect pacemakers. (See [\[link\]](#) and [\[link\]](#).) Physics allows you to understand the hazards of radiation and rationally evaluate these hazards more easily. Physics also explains the reason why a black car radiator helps remove heat in a car engine, and it explains why a white roof helps keep the inside of a house cool. Similarly, the operation of a car's ignition system as well as the transmission of electrical signals through our body's nervous system are

much easier to understand when you think about them in terms of basic physics.

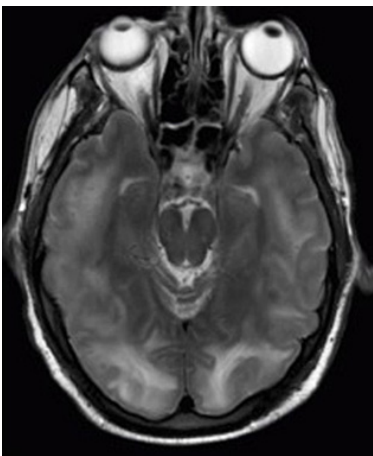
Physics is the foundation of many important disciplines and contributes directly to others. Chemistry, for example—since it deals with the interactions of atoms and molecules—is rooted in atomic and molecular physics. Most branches of engineering are applied physics. In architecture, physics is at the heart of structural stability, and is involved in the acoustics, heating, lighting, and cooling of buildings. Parts of geology rely heavily on physics, such as radioactive dating of rocks, earthquake analysis, and heat transfer in the Earth. Some disciplines, such as biophysics and geophysics, are hybrids of physics and other disciplines.

Physics has many applications in the biological sciences. On the microscopic level, it helps describe the properties of cell walls and cell membranes ([\[link\]](#) and [\[link\]](#)). On the macroscopic level, it can explain the heat, work, and power associated with the human body. Physics is involved in medical diagnostics, such as x-rays, magnetic resonance imaging (MRI), and ultrasonic blood flow measurements. Medical therapy sometimes directly involves physics; for example, cancer radiotherapy uses ionizing radiation. Physics can also explain sensory phenomena, such as how musical instruments make sound, how the eye detects color, and how lasers can transmit information.

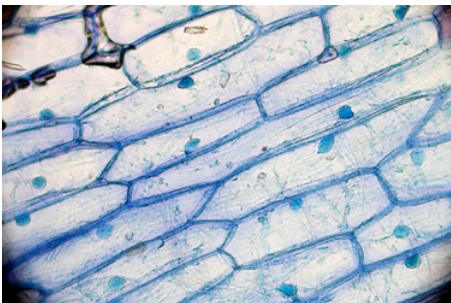
It is not necessary to formally study all applications of physics. What is most useful is knowledge of the basic laws of physics and a skill in the analytical methods for applying them. The study of physics also can improve your problem-solving skills. Furthermore, physics has retained the most basic aspects of science, so it is used by all of the sciences, and the study of physics makes other sciences easier to understand.



The laws of physics help us understand how common appliances work. For example, the laws of physics can help explain how microwave ovens heat up food, and they also help us understand why it is dangerous to place metal objects in a microwave oven. (credit: MoneyBlogNewz)

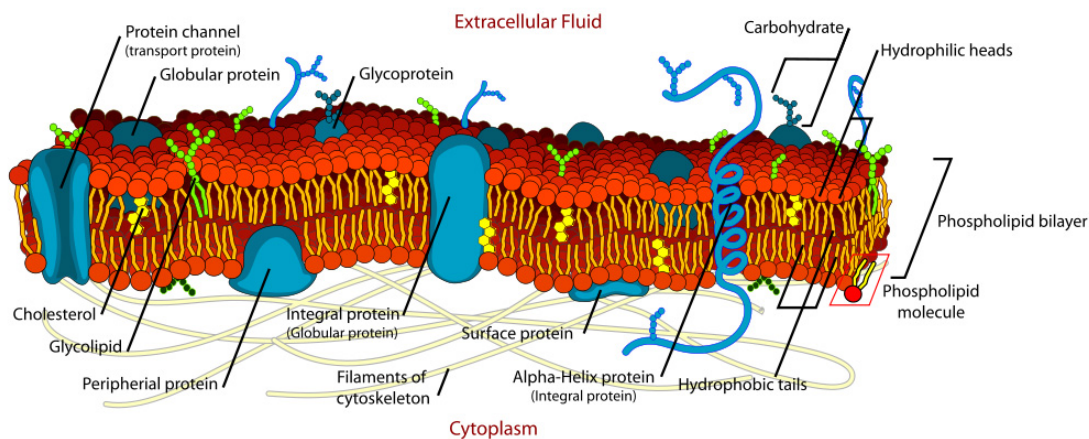


These two applications of physics have more in common than meets the eye. Microwave ovens use electromagnetic waves to heat food. Magnetic resonance imaging (MRI) also uses electromagnetic waves to yield an image of the brain, from which the exact location of tumors can be determined.
(credit: Rashmi Chawla, Daniel Smith, and Paul E. Marik)



Physics, chemistry,

and biology help describe the properties of cell walls in plant cells, such as the onion cells seen here. (credit: Umberto Salvagnin)

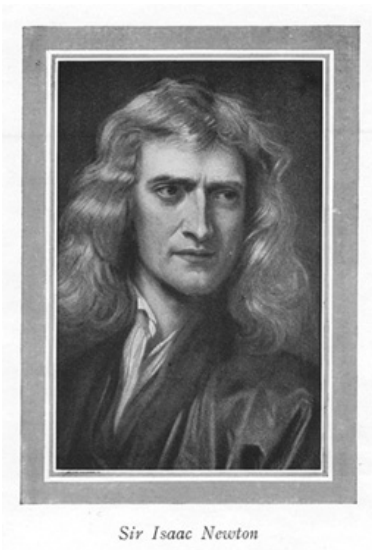


An artist's rendition of the the structure of a cell membrane. Membranes form the boundaries of animal cells and are complex in structure and function. Many of the most fundamental properties of life, such as the firing of nerve cells, are related to membranes. The disciplines of biology, chemistry, and physics all help us understand the membranes of animal cells. (credit: Mariana Ruiz)

Models, Theories, and Laws; The Role of Experimentation

The laws of nature are concise descriptions of the universe around us; they are human statements of the underlying laws or rules that all natural processes follow. Such laws are intrinsic to the universe; humans did not

create them and so cannot change them. We can only discover and understand them. Their discovery is a very human endeavor, with all the elements of mystery, imagination, struggle, triumph, and disappointment inherent in any creative effort. (See [\[link\]](#) and [\[link\]](#).) The cornerstone of discovering natural laws is observation; science must describe the universe as it is, not as we may imagine it to be.



Isaac Newton
(1642–1727) was very reluctant to publish his revolutionary work and had to be convinced to do so. In his later years, he stepped down from his academic post and became exchequer of the Royal Mint. He took this post

seriously,
inventing reeding
(or creating
ridges) on the
edge of coins to
prevent
unscrupulous
people from
trimming the
silver off of them
before using them
as currency.
(credit: Arthur
Shuster and
Arthur E. Shipley:
*Britain's Heritage
of Science*.
London, 1917.)



Marie Curie
(1867–1934)
sacrificed

monetary assets
to help finance
her early
research and
damaged her
physical well-
being with
radiation
exposure. She is
the only person
to win Nobel
prizes in both
physics and
chemistry. One
of her daughters
also won a
Nobel Prize.
(credit:
Wikimedia
Commons)

We all are curious to some extent. We look around, make generalizations, and try to understand what we see—for example, we look up and wonder whether one type of cloud signals an oncoming storm. As we become serious about exploring nature, we become more organized and formal in collecting and analyzing data. We attempt greater precision, perform controlled experiments (if we can), and write down ideas about how the data may be organized and unified. We then formulate models, theories, and laws based on the data we have collected and analyzed to generalize and communicate the results of these experiments.

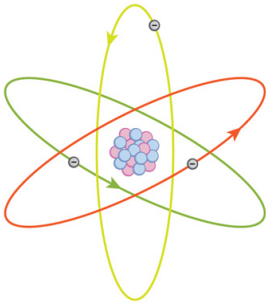
A **model** is a representation of something that is often too difficult (or impossible) to display directly. While a model is justified with experimental proof, it is only accurate under limited situations. An example is the planetary model of the atom in which electrons are pictured as orbiting the

nucleus, analogous to the way planets orbit the Sun. (See [\[link\]](#).) We cannot observe electron orbits directly, but the mental image helps explain the observations we can make, such as the emission of light from hot gases (atomic spectra). Physicists use models for a variety of purposes. For example, models can help physicists analyze a scenario and perform a calculation, or they can be used to represent a situation in the form of a computer simulation. A **theory** is an explanation for patterns in nature that is supported by scientific evidence and verified multiple times by various groups of researchers. Some theories include models to help visualize phenomena, whereas others do not. Newton's theory of gravity, for example, does not require a model or mental image, because we can observe the objects directly with our own senses. The kinetic theory of gases, on the other hand, is a model in which a gas is viewed as being composed of atoms and molecules. Atoms and molecules are too small to be observed directly with our senses—thus, we picture them mentally to understand what our instruments tell us about the behavior of gases.

A **law** uses concise language to describe a generalized pattern in nature that is supported by scientific evidence and repeated experiments. Often, a law can be expressed in the form of a single mathematical equation. Laws and theories are similar in that they are both scientific statements that result from a tested hypothesis and are supported by scientific evidence. However, the designation *law* is reserved for a concise and very general statement that describes phenomena in nature, such as the law that energy is conserved during any process, or Newton's second law of motion, which relates force, mass, and acceleration by the simple equation $\mathbf{F} = m\mathbf{a}$. A theory, in contrast, is a less concise statement of observed phenomena. For example, the Theory of Evolution and the Theory of Relativity cannot be expressed concisely enough to be considered a law. The biggest difference between a law and a theory is that a theory is much more complex and dynamic. A law describes a single action, whereas a theory explains an entire group of related phenomena. And, whereas a law is a postulate that forms the foundation of the scientific method, a theory is the end result of that process.

Less broadly applicable statements are usually called principles (such as Pascal's principle, which is applicable only in fluids), but the distinction

between laws and principles often is not carefully made.



What is a
model?

This
planetary
model of
the atom
shows
electrons
orbiting the
nucleus. It
is a
drawing
that we use
to form a
mental
image of
the atom
that we
cannot see
directly
with our
eyes
because it
is too
small.

Note:**Models, Theories, and Laws**

Models, theories, and laws are used to help scientists analyze the data they have already collected. However, often after a model, theory, or law has been developed, it points scientists toward new discoveries they would not otherwise have made.

The models, theories, and laws we devise sometimes *imply the existence of objects or phenomena as yet unobserved*. These predictions are remarkable triumphs and tributes to the power of science. It is the underlying order in the universe that enables scientists to make such spectacular predictions. However, if *experiment* does not verify our predictions, then the theory or law is wrong, no matter how elegant or convenient it is. Laws can never be known with absolute certainty because it is impossible to perform every imaginable experiment in order to confirm a law in every possible scenario. Physicists operate under the assumption that all scientific laws and theories are valid until a counterexample is observed. If a good-quality, verifiable experiment contradicts a well-established law, then the law must be modified or overthrown completely.

The study of science in general and physics in particular is an adventure much like the exploration of uncharted ocean. Discoveries are made; models, theories, and laws are formulated; and the beauty of the physical universe is made more sublime for the insights gained.

Note:**The Scientific Method**

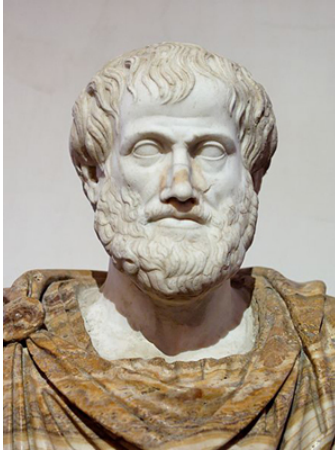
As scientists inquire and gather information about the world, they follow a process called the **scientific method**. This process typically begins with an observation and question that the scientist will research. Next, the scientist

typically performs some research about the topic and then devises a hypothesis. Then, the scientist will test the hypothesis by performing an experiment. Finally, the scientist analyzes the results of the experiment and draws a conclusion. Note that the scientific method can be applied to many situations that are not limited to science, and this method can be modified to suit the situation.

Consider an example. Let us say that you try to turn on your car, but it will not start. You undoubtedly wonder: Why will the car not start? You can follow a scientific method to answer this question. First off, you may perform some research to determine a variety of reasons why the car will not start. Next, you will state a hypothesis. For example, you may believe that the car is not starting because it has no engine oil. To test this, you open the hood of the car and examine the oil level. You observe that the oil is at an acceptable level, and you thus conclude that the oil level is not contributing to your car issue. To troubleshoot the issue further, you may devise a new hypothesis to test and then repeat the process again.

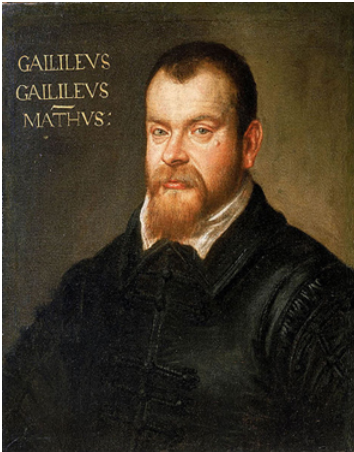
The Evolution of Natural Philosophy into Modern Physics

Physics was not always a separate and distinct discipline. It remains connected to other sciences to this day. The word *physics* comes from Greek, meaning nature. The study of nature came to be called “natural philosophy.” From ancient times through the Renaissance, natural philosophy encompassed many fields, including astronomy, biology, chemistry, physics, mathematics, and medicine. Over the last few centuries, the growth of knowledge has resulted in ever-increasing specialization and branching of natural philosophy into separate fields, with physics retaining the most basic facets. (See [\[link\]](#), [\[link\]](#), and [\[link\]](#).) Physics as it developed from the Renaissance to the end of the 19th century is called **classical physics**. It was transformed into modern physics by revolutionary discoveries made starting at the beginning of the 20th century.



Over the centuries, natural philosophy has evolved into more specialized disciplines, as illustrated by the contributions of some of the greatest minds in history. The Greek philosopher **Aristotle** (384–322 B.C.) wrote on a broad range of topics including physics, animals, the soul, politics, and poetry.
(credit: Jastrow

(2006)/Ludovisi
Collection)



Galileo Galilei
(1564–1642) laid
the foundation of
modern
experimentation
and made
contributions in
mathematics,
physics, and
astronomy.
(credit:
Domenico
Tintoretto)



Niels Bohr
(1885–1962)
made
fundamental
contributions to
the development
of quantum
mechanics, one
part of modern
physics. (credit:
United States
Library of
Congress Prints
and Photographs
Division)

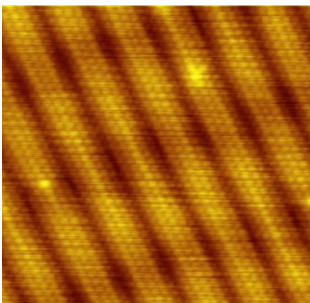
Classical physics is not an exact description of the universe, but it is an excellent approximation under the following conditions: Matter must be moving at speeds less than about 1% of the speed of light, the objects dealt with must be large enough to be seen with a microscope, and only weak gravitational fields, such as the field generated by the Earth, can be involved. Because humans live under such circumstances, classical physics seems intuitively reasonable, while many aspects of modern physics seem bizarre. This is why models are so useful in modern physics—they let us

conceptualize phenomena we do not ordinarily experience. We can relate to models in human terms and visualize what happens when objects move at high speeds or imagine what objects too small to observe with our senses might be like. For example, we can understand an atom's properties because we can picture it in our minds, although we have never seen an atom with our eyes. New tools, of course, allow us to better picture phenomena we cannot see. In fact, new instrumentation has allowed us in recent years to actually “picture” the atom.

Note:

Limits on the Laws of Classical Physics

For the laws of classical physics to apply, the following criteria must be met: Matter must be moving at speeds less than about 1% of the speed of light, the objects dealt with must be large enough to be seen with a microscope, and only weak gravitational fields (such as the field generated by the Earth) can be involved.



Using a
scanning
tunneling
microscope
(STM),
scientists can
see the
individual
atoms that

compose this
sheet of gold.
(credit:
Erwinrossen)

Some of the most spectacular advances in science have been made in modern physics. Many of the laws of classical physics have been modified or rejected, and revolutionary changes in technology, society, and our view of the universe have resulted. Like science fiction, modern physics is filled with fascinating objects beyond our normal experiences, but it has the advantage over science fiction of being very real. Why, then, is the majority of this text devoted to topics of classical physics? There are two main reasons: Classical physics gives an extremely accurate description of the universe under a wide range of everyday circumstances, and knowledge of classical physics is necessary to understand modern physics.

Modern physics itself consists of the two revolutionary theories, relativity and quantum mechanics. These theories deal with the very fast and the very small, respectively. **Relativity** must be used whenever an object is traveling at greater than about 1% of the speed of light or experiences a strong gravitational field such as that near the Sun. **Quantum mechanics** must be used for objects smaller than can be seen with a microscope. The combination of these two theories is *relativistic quantum mechanics*, and it describes the behavior of small objects traveling at high speeds or experiencing a strong gravitational field. Relativistic quantum mechanics is the best universally applicable theory we have. Because of its mathematical complexity, it is used only when necessary, and the other theories are used whenever they will produce sufficiently accurate results. We will find, however, that we can do a great deal of modern physics with the algebra and trigonometry used in this text.

Exercise:

Check Your Understanding

Problem:

A friend tells you he has learned about a new law of nature. What can you know about the information even before your friend describes the law? How would the information be different if your friend told you he had learned about a scientific theory rather than a law?

Solution:

Without knowing the details of the law, you can still infer that the information your friend has learned conforms to the requirements of all laws of nature: it will be a concise description of the universe around us; a statement of the underlying rules that all natural processes follow. If the information had been a theory, you would be able to infer that the information will be a large-scale, broadly applicable generalization.

Note:

PhET Explorations: Equation Grapher

Learn about graphing polynomials. The shape of the curve changes as the constants are adjusted. View the curves for the individual terms (e.g. $y = bx$) to see how they add to generate the polynomial curve.

https://phet.colorado.edu/sims/equation-grapher/equation-grapher_en.html

Summary

- Science seeks to discover and describe the underlying order and simplicity in nature.
- Physics is the most basic of the sciences, concerning itself with energy, matter, space and time, and their interactions.
- Scientific laws and theories express the general truths of nature and the body of knowledge they encompass. These laws of nature are rules that all natural processes appear to follow.

Conceptual Questions

Exercise:

Problem:

Models are particularly useful in relativity and quantum mechanics, where conditions are outside those normally encountered by humans. What is a model?

Exercise:

Problem: How does a model differ from a theory?

Exercise:

Problem:

If two different theories describe experimental observations equally well, can one be said to be more valid than the other (assuming both use accepted rules of logic)?

Exercise:

Problem: What determines the validity of a theory?

Exercise:

Problem:

Certain criteria must be satisfied if a measurement or observation is to be believed. Will the criteria necessarily be as strict for an expected result as for an unexpected result?

Exercise:

Problem:

Can the validity of a model be limited, or must it be universally valid? How does this compare to the required validity of a theory or a law?

Exercise:

Problem:

Classical physics is a good approximation to modern physics under certain circumstances. What are they?

Exercise:

Problem: When is it *necessary* to use relativistic quantum mechanics?

Exercise:**Problem:**

Can classical physics be used to accurately describe a satellite moving at a speed of 7500 m/s? Explain why or why not.

Glossary

classical physics

physics that was developed from the Renaissance to the end of the 19th century

physics

the science concerned with describing the interactions of energy, matter, space, and time; it is especially interested in what fundamental mechanisms underlie every phenomenon

model

representation of something that is often too difficult (or impossible) to display directly

theory

an explanation for patterns in nature that is supported by scientific evidence and verified multiple times by various groups of researchers

law

a description, using concise language or a mathematical formula, a generalized pattern in nature that is supported by scientific evidence

and repeated experiments

scientific method

a method that typically begins with an observation and question that the scientist will research; next, the scientist typically performs some research about the topic and then devises a hypothesis; then, the scientist will test the hypothesis by performing an experiment; finally, the scientist analyzes the results of the experiment and draws a conclusion

modern physics

the study of relativity, quantum mechanics, or both

relativity

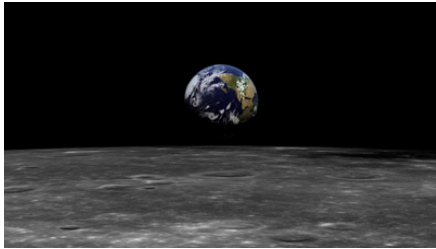
the study of objects moving at speeds greater than about 1% of the speed of light, or of objects being affected by a strong gravitational field

quantum mechanics

the study of objects smaller than can be seen with a microscope

Physical Quantities and Units

- Perform unit conversions both in the SI and English units.
- Explain the most common prefixes in the SI units and be able to write them in scientific notation.

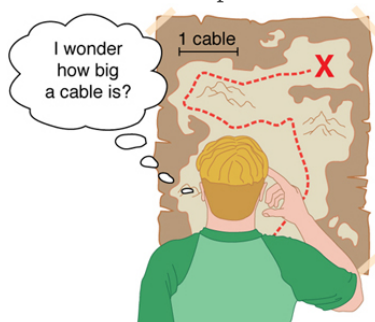


The distance from Earth to the Moon may seem immense, but it is just a tiny fraction of the distances from Earth to other celestial bodies. (credit: NASA)

The range of objects and phenomena studied in physics is immense. From the incredibly short lifetime of a nucleus to the age of the Earth, from the tiny sizes of sub-nuclear particles to the vast distance to the edges of the known universe, from the force exerted by a jumping flea to the force between Earth and the Sun, there are enough factors of 10 to challenge the imagination of even the most experienced scientist. Giving numerical values for physical quantities and equations for physical principles allows us to understand nature much more deeply than does qualitative description alone. To comprehend these vast ranges, we must also have accepted units in which to express them. And we shall find that (even in the potentially mundane discussion of meters, kilograms, and seconds) a profound simplicity of nature appears—all physical quantities can be expressed as combinations of only four fundamental physical quantities: length, mass, time, and electric current.

We define a **physical quantity** either by *specifying how it is measured* or by *stating how it is calculated* from other measurements. For example, we define distance and time by specifying methods for measuring them, whereas we define *average speed* by stating that it is calculated as distance traveled divided by time of travel.

Measurements of physical quantities are expressed in terms of **units**, which are standardized values. For example, the length of a race, which is a physical quantity, can be expressed in units of meters (for sprinters) or kilometers (for distance runners). Without standardized units, it would be extremely difficult for scientists to express and compare measured values in a meaningful way. (See [\[link\]](#).)



Distances given in
unknown units are
maddeningly useless.

There are two major systems of units used in the world: **SI units** (also known as the metric system) and **English units** (also known as the customary or imperial system). **English units** were historically used in nations once ruled by the British Empire and are still widely used in the United States. Virtually every other country in the world now uses SI units as the standard; the metric system is also the standard system agreed upon by scientists and mathematicians. The acronym “SI” is derived from the French *Système International*.

SI Units: Fundamental and Derived Units

[\[link\]](#) gives the fundamental SI units that are used throughout this textbook. This text uses non-SI units in a few applications where they are in very common use, such as the measurement of blood pressure in millimeters of mercury (mm Hg). Whenever non-SI units are discussed, they will be tied to SI units through conversions.

Length	Mass	Time	Electric Current
meter (m)	kilogram (kg)	second (s)	ampere (A)

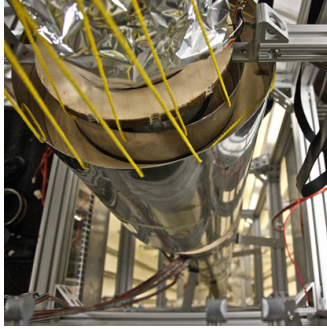
Fundamental SI Units

It is an intriguing fact that some physical quantities are more fundamental than others and that the most fundamental physical quantities can be defined *only* in terms of the procedure used to measure them. The units in which they are measured are thus called **fundamental units**. In this textbook, the fundamental physical quantities are taken to be length, mass, time, and electric current. (Note that electric current will not be introduced until much later in this text.) All other physical quantities, such as force and electric charge, can be expressed as algebraic combinations of length, mass, time, and current (for example, speed is length divided by time); these units are called **derived units**.

Units of Time, Length, and Mass: The Second, Meter, and Kilogram

The Second

The SI unit for time, the **second**(abbreviated s), has a long history. For many years it was defined as 1/86,400 of a mean solar day. More recently, a new standard was adopted to gain greater accuracy and to define the second in terms of a non-varying, or constant, physical phenomenon (because the solar day is getting longer due to very gradual slowing of the Earth’s rotation). Cesium atoms can be made to vibrate in a very steady way, and these vibrations can be readily observed and counted. In 1967 the second was redefined as the time required for 9,192,631,770 of these vibrations. (See [\[link\]](#).) Accuracy in the fundamental units is essential, because all measurements are ultimately expressed in terms of fundamental units and can be no more accurate than are the fundamental units themselves.



An atomic clock such as this one uses the vibrations of cesium atoms to keep time to a precision of better than a microsecond per year. The fundamental unit of time, the second, is based on such clocks. This image is looking down from the top of an atomic fountain nearly 30 feet tall!
(credit: Steve Jurvetson/Flickr)

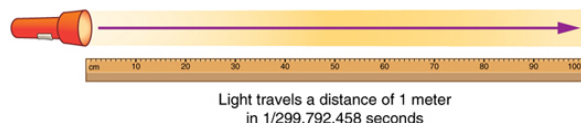
The Meter

The SI unit for length is the **meter** (abbreviated m); its definition has also changed over time to become more accurate and precise. The meter was first defined in 1791 as 1/10,000,000 of the distance from the equator to the North Pole. This measurement was improved in 1889 by redefining the meter to be the distance between two engraved lines on a platinum-iridium bar now kept near Paris. By 1960, it had become possible to define the meter even more accurately in terms of the wavelength of light, so it was again redefined as 1,650,763.73 wavelengths of orange light emitted by krypton atoms. In 1983, the meter was given its present definition (partly for greater accuracy) as the distance light travels in a vacuum in 1/299,792,458 of a second. (See [\[link\]](#).) This change defines the speed of light to be exactly 299,792,458 meters per second. The length of the meter will change if the speed of light is someday measured with greater accuracy.

The Kilogram

The SI unit for mass is the **kilogram** (abbreviated kg); it is defined to be the mass of a platinum-iridium cylinder kept with the old meter standard at the International Bureau of Weights and Measures near Paris. Exact replicas of the standard kilogram are also kept at the United States' National Institute of Standards

and Technology, or NIST, located in Gaithersburg, Maryland outside of Washington D.C., and at other locations around the world. The determination of all other masses can be ultimately traced to a comparison with the standard mass.



The meter is defined to be the distance light travels in $1/299,792,458$ of a second in a vacuum. Distance traveled is speed multiplied by time.

Electric current and its accompanying unit, the ampere, will be introduced in [Introduction to Electric Current, Resistance, and Ohm's Law](#) when electricity and magnetism are covered. The initial modules in this textbook are concerned with mechanics, fluids, heat, and waves. In these subjects all pertinent physical quantities can be expressed in terms of the fundamental units of length, mass, and time.

Metric Prefixes

SI units are part of the **metric system**. The metric system is convenient for scientific and engineering calculations because the units are categorized by factors of 10. [\[link\]](#) gives metric prefixes and symbols used to denote various factors of 10.

Metric systems have the advantage that conversions of units involve only powers of 10. There are 100 centimeters in a meter, 1000 meters in a kilometer, and so on. In nonmetric systems, such as the system of U.S. customary units, the relationships are not as simple—there are 12 inches in a foot, 5280 feet in a mile, and so on. Another advantage of the metric system is that the same unit can be used over extremely large ranges of values simply by using an appropriate metric prefix. For example, distances in meters are suitable in construction, while distances in kilometers are appropriate for air travel, and the tiny measure of nanometers are convenient in optical design. With the metric system there is no need to invent new units for particular applications.

The term **order of magnitude** refers to the scale of a value expressed in the metric system. Each power of 10 in the metric system represents a different order of magnitude. For example, 10^1 , 10^2 , 10^3 , and so forth are all different orders of magnitude. All quantities that can be expressed as a product of a specific power of 10 are said to be of the *same* order of magnitude. For example, the number 800 can be written as 8×10^2 , and the number 450 can be written as 4.5×10^2 . Thus, the numbers 800 and 450 are of the same order of magnitude: 10^2 . Order of magnitude can be thought of as a ballpark estimate for the scale of a value. The diameter of an atom is on the order of 10^{-9} m, while the diameter of the Sun is on the order of 10^9 m.

Note:

The Quest for Microscopic Standards for Basic Units

The fundamental units described in this chapter are those that produce the greatest accuracy and precision in measurement. There is a sense among physicists that, because there is an underlying microscopic substructure to matter, it would be most satisfying to base our standards of measurement on microscopic objects and fundamental physical phenomena such as the speed of light. A microscopic standard has been accomplished for the standard of time, which is based on the oscillations of the cesium atom.

The standard for length was once based on the wavelength of light (a small-scale length) emitted by a certain type of atom, but it has been supplanted by the more precise measurement of the speed of light. If it becomes possible to measure the mass of atoms or a particular arrangement of atoms such as a silicon sphere to greater precision than the kilogram standard, it may become possible to base mass measurements on the small scale. There are also possibilities that electrical phenomena on the small scale may someday allow us to base a unit of charge on the charge of electrons and protons, but at present current and charge are related to large-scale currents and forces between wires.

Prefix	Symbol	Value ^[footnote] See Appendix A for a discussion of powers of 10.	Example (some are approximate)			
exa	E	10^{18}	exameter	Em	10^{18} m	distance light travels in a century
peta	P	10^{15}	petasecond	Ps	10^{15} s	30 million years
tera	T	10^{12}	terawatt	TW	10^{12} W	powerful laser output
giga	G	10^9	gigahertz	GHz	10^9 Hz	a microwave frequency
mega	M	10^6	megacurie	MCi	10^6 Ci	high radioactivity
kilo	k	10^3	kilometer	km	10^3 m	about 6/10 mile
hecto	h	10^2	hectoliter	hL	10^2 L	26 gallons

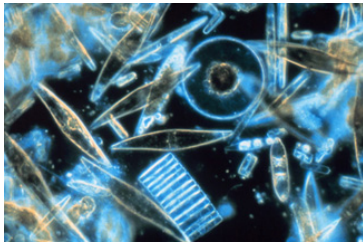
Prefix	Symbol	Value ^{[footnote]} See Appendix A for a discussion of powers of 10.	Example (some are approximate)			
deka	da	10^1	dekagram	dag	10^1 g	teaspoon of butter
—	—	10^0 (=1)				
deci	d	10^{-1}	deciliter	dL	10^{-1} L	less than half a soda
centi	c	10^{-2}	centimeter	cm	10^{-2} m	fingertip thickness
milli	m	10^{-3}	millimeter	mm	10^{-3} m	flea at its shoulders
micro	μ	10^{-6}	micrometer	μm	10^{-6} m	detail in microscope
nano	n	10^{-9}	nanogram	ng	10^{-9} g	small speck of dust
pico	p	10^{-12}	picofarad	pF	10^{-12} F	small capacitor in radio
femto	f	10^{-15}	femtometer	fm	10^{-15} m	size of a proton
atto	a	10^{-18}	attosecond	as	10^{-18} s	time light crosses an atom

Metric Prefixes for Powers of 10 and their Symbols

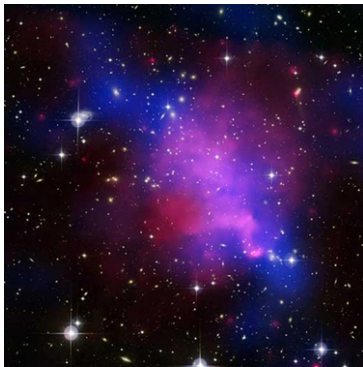
Known Ranges of Length, Mass, and Time

The vastness of the universe and the breadth over which physics applies are illustrated by the wide range of examples of known lengths, masses, and times in [\[link\]](#). Examination of this table will give you some

feeling for the range of possible topics and numerical values. (See [\[link\]](#) and [\[link\]](#).)



Tiny phytoplankton
swims among crystals of
ice in the Antarctic Sea.
They range from a few
micrometers to as much
as 2 millimeters in length.
(credit: Prof. Gordon T.
Taylor, Stony Brook
University; NOAA Corps
Collections)



Galaxies collide 2.4
billion light years away
from Earth. The
tremendous range of
observable phenomena in
nature challenges the
imagination. (credit:
NASA/CXC/UVic./A.
Mahdavi et al.
Optical/lensing:
CFHT/UVic./H. Hoekstra
et al.)

Unit Conversion and Dimensional Analysis

It is often necessary to convert from one type of unit to another. For example, if you are reading a European cookbook, some quantities may be expressed in units of liters and you need to convert them to cups. Or, perhaps you are reading walking directions from one location to another and you are interested in how many miles you will be walking. In this case, you will need to convert units of feet to miles.

Let us consider a simple example of how to convert units. Let us say that we want to convert 80 meters (m) to kilometers (km).

The first thing to do is to list the units that you have and the units that you want to convert to. In this case, we have units in *meters* and we want to convert to *kilometers*.

Next, we need to determine a **conversion factor** relating meters to kilometers. A conversion factor is a ratio expressing how many of one unit are equal to another unit. For example, there are 12 inches in 1 foot, 100 centimeters in 1 meter, 60 seconds in 1 minute, and so on. In this case, we know that there are 1,000 meters in 1 kilometer.

Now we can set up our unit conversion. We will write the units that we have and then multiply them by the conversion factor so that the units cancel out, as shown:

Equation:

80 m × (1 km / 1000 m) = 0.080 km.

Note that the unwanted m unit cancels, leaving only the desired km unit. You can use this method to convert between any types of unit.

Click [\[link\]](#) for a more complete list of conversion factors.

Lengths in meters		Masses in kilograms (more precise values in parentheses)		Times in seconds (more precise values in parentheses)	
10 ⁻¹⁸	Present experimental limit to smallest observable detail	10 ⁻³⁰	Mass of an electron (9.11 × 10 ⁻³¹ kg)	10 ⁻²³	Time for light to cross a proton
10 ⁻¹⁵	Diameter of a proton	10 ⁻²⁷	Mass of a hydrogen atom (1.67 × 10 ⁻²⁷ kg)	10 ⁻²²	Mean life of an extremely unstable nucleus

Lengths in meters		Masses in kilograms (more precise values in parentheses)		Times in seconds (more precise values in parentheses)	
10^{-14}	Diameter of a uranium nucleus	10^{-15}	Mass of a bacterium	10^{-15}	Time for one oscillation of visible light
10^{-10}	Diameter of a hydrogen atom	10^{-5}	Mass of a mosquito	10^{-13}	Time for one vibration of an atom in a solid
10^{-8}	Thickness of membranes in cells of living organisms	10^{-2}	Mass of a hummingbird	10^{-8}	Time for one oscillation of an FM radio wave
10^{-6}	Wavelength of visible light	1	Mass of a liter of water (about a quart)	10^{-3}	Duration of a nerve impulse
10^{-3}	Size of a grain of sand	10^2	Mass of a person	1	Time for one heartbeat
1	Height of a 4-year-old child	10^3	Mass of a car	10^5	One day (8.64×10^4 s)
10^2	Length of a football field	10^8	Mass of a large ship	10^7	One year (y) (3.16×10^7 s)
10^4	Greatest ocean depth	10^{12}	Mass of a large iceberg	10^9	About half the life expectancy of a human
10^7	Diameter of the Earth	10^{15}	Mass of the nucleus of a comet	10^{11}	Recorded history
10^{11}	Distance from the Earth to the Sun	10^{23}	Mass of the Moon (7.35×10^{22} kg)	10^{17}	Age of the Earth
10^{16}	Distance traveled by light in 1 year (a light year)	10^{25}	Mass of the Earth (5.97×10^{24} kg)	10^{18}	Age of the universe
10^{21}	Diameter of the Milky Way galaxy	10^{30}	Mass of the Sun (1.99×10^{30} kg)		

Lengths in meters		Masses in kilograms (more precise values in parentheses)		Times in seconds (more precise values in parentheses)	
10^{22}	Distance from the Earth to the nearest large galaxy (Andromeda)	10^{42}	Mass of the Milky Way galaxy (current upper limit)		
10^{26}	Distance from the Earth to the edges of the known universe	10^{53}	Mass of the known universe (current upper limit)		

Approximate Values of Length, Mass, and Time

Example:

Unit Conversions: A Short Drive Home

Suppose that you drive the 10.0 km from your university to home in 20.0 min. Calculate your average speed (a) in kilometers per hour (km/h) and (b) in meters per second (m/s). (Note: Average speed is distance traveled divided by time of travel.)

Strategy

First we calculate the average speed using the given units. Then we can get the average speed into the desired units by picking the correct conversion factor and multiplying by it. The correct conversion factor is the one that cancels the unwanted unit and leaves the desired unit in its place.

Solution for (a)

(1) Calculate average speed. Average speed is distance traveled divided by time of travel. (Take this definition as a given for now—average speed and other motion concepts will be covered in a later module.) In equation form,

Equation:

$$\text{average speed} = \frac{\text{distance}}{\text{time}}.$$

(2) Substitute the given values for distance and time.

Equation:

$$\text{average speed} = \frac{10.0 \text{ km}}{20.0 \text{ min}} = 0.500 \frac{\text{km}}{\text{min}}.$$

(3) Convert km/min to km/h: multiply by the conversion factor that will cancel minutes and leave hours. That conversion factor is 60 min/hr. Thus,

Equation:

$$\text{average speed} = 0.500 \frac{\text{km}}{\text{min}} \times \frac{60 \text{ min}}{1 \text{ h}} = 30.0 \frac{\text{km}}{\text{h}}.$$

Discussion for (a)

To check your answer, consider the following:

(1) Be sure that you have properly cancelled the units in the unit conversion. If you have written the unit conversion factor upside down, the units will not cancel properly in the equation. If you accidentally get the ratio upside down, then the units will not cancel; rather, they will give you the wrong units as follows:

Equation:

$$\frac{\text{km}}{\text{min}} \times \frac{1 \text{ hr}}{60 \text{ min}} = \frac{1}{60} \frac{\text{km} \cdot \text{hr}}{\text{min}^2},$$

which are obviously not the desired units of km/h.

(2) Check that the units of the final answer are the desired units. The problem asked us to solve for average speed in units of km/h and we have indeed obtained these units.

(3) Check the significant figures. Because each of the values given in the problem has three significant figures, the answer should also have three significant figures. The answer 30.0 km/hr does indeed have three significant figures, so this is appropriate. Note that the significant figures in the conversion factor are not relevant because an hour is *defined* to be 60 minutes, so the precision of the conversion factor is perfect.

(4) Next, check whether the answer is reasonable. Let us consider some information from the problem—if you travel 10 km in a third of an hour (20 min), you would travel three times that far in an hour. The answer does seem reasonable.

Solution for (b)

There are several ways to convert the average speed into meters per second.

(1) Start with the answer to (a) and convert km/h to m/s. Two conversion factors are needed—one to convert hours to seconds, and another to convert kilometers to meters.

(2) Multiplying by these yields

Equation:

$$\text{Average speed} = 30.0 \frac{\text{km}}{\text{h}} \times \frac{1 \text{ h}}{3,600 \text{ s}} \times \frac{1,000 \text{ m}}{1 \text{ km}},$$

Equation:

$$\text{Average speed} = 8.33 \frac{\text{m}}{\text{s}}.$$

Discussion for (b)

If we had started with 0.500 km/min, we would have needed different conversion factors, but the answer would have been the same: 8.33 m/s.

You may have noted that the answers in the worked example just covered were given to three digits.

Why? When do you need to be concerned about the number of digits in something you calculate? Why not write down all the digits your calculator produces? The module [Accuracy, Precision, and Significant Figures](#) will help you answer these questions.

Note:

Nonstandard Units

While there are numerous types of units that we are all familiar with, there are others that are much more obscure. For example, a **firkin** is a unit of volume that was once used to measure beer. One firkin equals about 34 liters. To learn more about nonstandard units, use a dictionary or encyclopedia to research different “weights and measures.” Take note of any unusual units, such as a barleycorn, that are not listed in the text. Think about how the unit is defined and state its relationship to SI units.

Exercise:
Check Your Understanding

Problem:

Some hummingbirds beat their wings more than 50 times per second. A scientist is measuring the time it takes for a hummingbird to beat its wings once. Which fundamental unit should the scientist use to describe the measurement? Which factor of 10 is the scientist likely to use to describe the motion precisely? Identify the metric prefix that corresponds to this factor of 10.

Solution:

The scientist will measure the time between each movement using the fundamental unit of seconds. Because the wings beat so fast, the scientist will probably need to measure in milliseconds, or 10^{-3} seconds. (50 beats per second corresponds to 20 milliseconds per beat.)

Exercise:
Check Your Understanding

Problem:

One cubic centimeter is equal to one milliliter. What does this tell you about the different units in the SI metric system?

Solution:

The fundamental unit of length (meter) is probably used to create the derived unit of volume (liter). The measure of a milliliter is dependent on the measure of a centimeter.

Summary

- Physical quantities are a characteristic or property of an object that can be measured or calculated from other measurements.
- Units are standards for expressing and comparing the measurement of physical quantities. All units can be expressed as combinations of four fundamental units.
- The four fundamental units we will use in this text are the meter (for length), the kilogram (for mass), the second (for time), and the ampere (for electric current). These units are part of the metric system, which uses powers of 10 to relate quantities over the vast ranges encountered in nature.
- The four fundamental units are abbreviated as follows: meter, m; kilogram, kg; second, s; and ampere, A. The metric system also uses a standard set of prefixes to denote each order of magnitude greater than or lesser than the fundamental unit itself.
- Unit conversions involve changing a value expressed in one type of unit to another type of unit. This is done by using conversion factors, which are ratios relating equal quantities of different units.

Conceptual Questions

Exercise:

Problem: Identify some advantages of metric units.

Problems & Exercises

Exercise:**Problem:**

The speed limit on some interstate highways is roughly 100 km/h. (a) What is this in meters per second? (b) How many miles per hour is this?

Solution:

- a. 27.8 m/s
- b. 62.1 mph

Exercise:**Problem:**

A car is traveling at a speed of 33 m/s. (a) What is its speed in kilometers per hour? (b) Is it exceeding the 90 km/h speed limit?

Exercise:**Problem:**

Show that $1.0 \text{ m/s} = 3.6 \text{ km/h}$. Hint: Show the explicit steps involved in converting $1.0 \text{ m/s} = 3.6 \text{ km/h}$.

Solution:

$$\begin{aligned}\frac{1.0 \text{ m}}{\text{s}} &= \frac{1.0 \text{ m}}{\text{s}} \times \frac{3600 \text{ s}}{1 \text{ hr}} \times \frac{1 \text{ km}}{1000 \text{ m}} \\ &= 3.6 \text{ km/h.}\end{aligned}$$

Exercise:**Problem:**

American football is played on a 100-yd-long field, excluding the end zones. How long is the field in meters? (Assume that 1 meter equals 3.281 feet.)

Exercise:**Problem:**

Soccer fields vary in size. A large soccer field is 115 m long and 85 m wide. What are its dimensions in feet and inches? (Assume that 1 meter equals 3.281 feet.)

Solution:

length: 377 ft; 4.53×10^3 in. width: 280 ft; 3.3×10^3 in.

Exercise:**Problem:**

What is the height in meters of a person who is 6 ft 1.0 in. tall? (Assume that 1 meter equals 39.37 in.)

Exercise:

Problem:

Mount Everest, at 29,028 feet, is the tallest mountain on the Earth. What is its height in kilometers? (Assume that 1 kilometer equals 3,281 feet.)

Solution:

8.847 km

Exercise:

Problem: The speed of sound is measured to be 342 m/s on a certain day. What is this in km/h?

Exercise:**Problem:**

Tectonic plates are large segments of the Earth's crust that move slowly. Suppose that one such plate has an average speed of 4.0 cm/year. (a) What distance does it move in 1 s at this speed? (b) What is its speed in kilometers per million years?

Solution:

(a) 1.3×10^{-9} m

(b) 40 km/My

Exercise:**Problem:**

(a) Refer to [\[link\]](#) to determine the average distance between the Earth and the Sun. Then calculate the average speed of the Earth in its orbit in kilometers per second. (b) What is this in meters per second?

Glossary

physical quantity

a characteristic or property of an object that can be measured or calculated from other measurements

units

a standard used for expressing and comparing measurements

SI units

the international system of units that scientists in most countries have agreed to use; includes units such as meters, liters, and grams

English units

system of measurement used in the United States; includes units of measurement such as feet, gallons, and pounds

fundamental units

units that can only be expressed relative to the procedure used to measure them

derived units

units that can be calculated using algebraic combinations of the fundamental units

second

the SI unit for time, abbreviated (s)

meter

the SI unit for length, abbreviated (m)

kilogram

the SI unit for mass, abbreviated (kg)

metric system

a system in which values can be calculated in factors of 10

order of magnitude

refers to the size of a quantity as it relates to a power of 10

conversion factor

a ratio expressing how many of one unit are equal to another unit

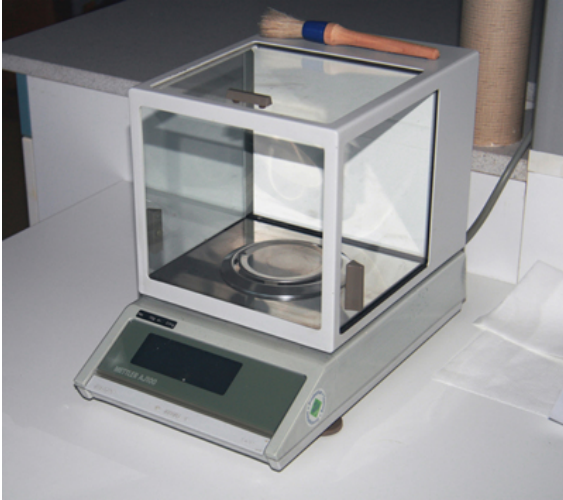
Accuracy, Precision, and Significant Figures

- Determine the appropriate number of significant figures in both addition and subtraction, as well as multiplication and division calculations.
- Calculate the percent uncertainty of a measurement.



A double-pan mechanical balance is used to compare different masses. Usually an object with unknown mass is placed in one pan and objects of known mass are placed in the other pan. When the bar that connects the two pans is horizontal, then the masses in both pans are equal. The “known masses” are typically metal cylinders of standard mass such as 1 gram, 10 grams, and 100 grams.

(credit: Serge Melki)



Many mechanical balances, such as double-pan balances, have been replaced by digital scales, which can typically measure the mass of an object more precisely. Whereas a mechanical balance may only read the mass of an object to the nearest tenth of a gram, many digital scales can measure the mass of an object up to the nearest thousandth of a gram. (credit: Karel Jakubec)

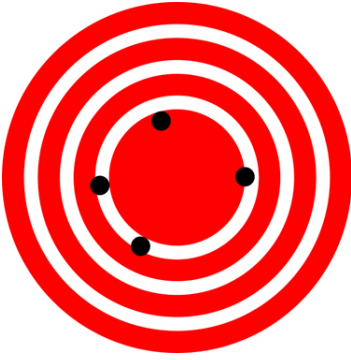
Accuracy and Precision of a Measurement

Science is based on observation and experiment—that is, on measurements. **Accuracy** is how close a measurement is to the correct value for that measurement. For example, let us say that you are measuring the length of standard computer paper. The packaging in which you purchased the paper states that it is 11.0 inches long. You measure the length of the paper three times and obtain the following measurements: 11.1 in., 11.2 in., and 10.9 in.

These measurements are quite accurate because they are very close to the correct value of 11.0 inches. In contrast, if you had obtained a measurement of 12 inches, your measurement would not be very accurate.

The **precision** of a measurement system refers to how close the agreement is between repeated measurements (which are repeated under the same conditions). Consider the example of the paper measurements. The precision of the measurements refers to the spread of the measured values. One way to analyze the precision of the measurements would be to determine the range, or difference, between the lowest and the highest measured values. In that case, the lowest value was 10.9 in. and the highest value was 11.2 in. Thus, the measured values deviated from each other by at most 0.3 in. These measurements were relatively precise because they did not vary too much in value. However, if the measured values had been 10.9, 11.1, and 11.9, then the measurements would not be very precise because there would be significant variation from one measurement to another.

The measurements in the paper example are both accurate and precise, but in some cases, measurements are accurate but not precise, or they are precise but not accurate. Let us consider an example of a GPS system that is attempting to locate the position of a restaurant in a city. Think of the restaurant location as existing at the center of a bull's-eye target, and think of each GPS attempt to locate the restaurant as a black dot. In [\[link\]](#), you can see that the GPS measurements are spread out far apart from each other, but they are all relatively close to the actual location of the restaurant at the center of the target. This indicates a low precision, high accuracy measuring system. However, in [\[link\]](#), the GPS measurements are concentrated quite closely to one another, but they are far away from the target location. This indicates a high precision, low accuracy measuring system.



A GPS system attempts to locate a restaurant at the center of the bull's-eye. The black dots represent each attempt to pinpoint the location of the restaurant. The dots are spread out quite far apart from one another, indicating low precision, but they are each rather close to the actual location of the restaurant, indicating high accuracy.
(credit: Dark Evil)



In this figure,
the dots are
concentrated
rather closely to
one another,
indicating high
precision, but
they are rather
far away from
the actual
location of the
restaurant,
indicating low
accuracy.
(credit: Dark
Evil)

Accuracy, Precision, and Uncertainty

The degree of accuracy and precision of a measuring system are related to the **uncertainty** in the measurements. Uncertainty is a quantitative measure of how much your measured values deviate from a standard or expected value. If your measurements are not very accurate or precise, then the

uncertainty of your values will be very high. In more general terms, uncertainty can be thought of as a disclaimer for your measured values. For example, if someone asked you to provide the mileage on your car, you might say that it is 45,000 miles, plus or minus 500 miles. The plus or minus amount is the uncertainty in your value. That is, you are indicating that the actual mileage of your car might be as low as 44,500 miles or as high as 45,500 miles, or anywhere in between. All measurements contain some amount of uncertainty. In our example of measuring the length of the paper, we might say that the length of the paper is 11 in., plus or minus 0.2 in. The uncertainty in a measurement, A , is often denoted as δA (“delta A ”), so the measurement result would be recorded as $A \pm \delta A$. In our paper example, the length of the paper could be expressed as $11 \text{ in.} \pm 0.2$.

The factors contributing to uncertainty in a measurement include:

1. Limitations of the measuring device,
2. The skill of the person making the measurement,
3. Irregularities in the object being measured,
4. Any other factors that affect the outcome (highly dependent on the situation).

In our example, such factors contributing to the uncertainty could be the following: the smallest division on the ruler is 0.1 in., the person using the ruler has bad eyesight, or one side of the paper is slightly longer than the other. At any rate, the uncertainty in a measurement must be based on a careful consideration of all the factors that might contribute and their possible effects.

Note:

Making Connections: Real-World Connections – Fevers or Chills?

Uncertainty is a critical piece of information, both in physics and in many other real-world applications. Imagine you are caring for a sick child. You suspect the child has a fever, so you check his or her temperature with a thermometer. What if the uncertainty of the thermometer were 3.0°C ? If the child’s temperature reading was 37.0°C (which is normal body temperature), the “true” temperature could be anywhere from a

hypothermic 34.0°C to a dangerously high 40.0°C. A thermometer with an uncertainty of 3.0°C would be useless.

Percent Uncertainty

One method of expressing uncertainty is as a percent of the measured value. If a measurement A is expressed with uncertainty, δA , the **percent uncertainty** (%unc) is defined to be

Equation:

$$\% \text{ unc} = \frac{\delta A}{A} \times 100\%.$$

Example:

Calculating Percent Uncertainty: A Bag of Apples

A grocery store sells 5-lb bags of apples. You purchase four bags over the course of a month and weigh the apples each time. You obtain the following measurements:

Week 1 weight: 4.8 lb

Week 2 weight: 5.3 lb

Week 3 weight: 4.9 lb

Week 4 weight: 5.4 lb

You determine that the weight of the 5-lb bag has an uncertainty of ± 0.4 lb. What is the percent uncertainty of the bag's weight?

Strategy

First, observe that the expected value of the bag's weight, A , is 5 lb. The uncertainty in this value, δA , is 0.4 lb. We can use the following equation to determine the percent uncertainty of the weight:

Equation:

$$\% \text{ unc} = \frac{\delta A}{A} \times 100\%.$$

Solution

Plug the known values into the equation:

Equation:

$$\% \text{ unc} = \frac{0.4 \text{ lb}}{5 \text{ lb}} \times 100\% = 8\%.$$

Discussion

We can conclude that the weight of the apple bag is $5 \text{ lb} \pm 8\%$. Consider how this percent uncertainty would change if the bag of apples were half as heavy, but the uncertainty in the weight remained the same. Hint for future calculations: when calculating percent uncertainty, always remember that you must multiply the fraction by 100%. If you do not do this, you will have a decimal quantity, not a percent value.

Uncertainties in Calculations

There is an uncertainty in anything calculated from measured quantities. For example, the area of a floor calculated from measurements of its length and width has an uncertainty because the length and width have uncertainties. How big is the uncertainty in something you calculate by multiplication or division? If the measurements going into the calculation have small uncertainties (a few percent or less), then the **method of adding percents** can be used for multiplication or division. This method says that *the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation*. For example, if a floor has a length of 4.00 m and a width of 3.00 m, with uncertainties of 2% and 1%, respectively, then the area of the floor is 12.0 m^2 and has an uncertainty of 3%. (Expressed as an area this is 0.36 m^2 , which we round to 0.4 m^2 since the area of the floor is given to a tenth of a square meter.)

Exercise:

Check Your Understanding

Problem:

A high school track coach has just purchased a new stopwatch. The stopwatch manual states that the stopwatch has an uncertainty of ± 0.05 s. Runners on the track coach's team regularly clock 100-m sprints of 11.49 s to 15.01 s. At the school's last track meet, the first-place sprinter came in at 12.04 s and the second-place sprinter came in at 12.07 s. Will the coach's new stopwatch be helpful in timing the sprint team? Why or why not?

Solution:

No, the uncertainty in the stopwatch is too great to effectively differentiate between the sprint times.

Precision of Measuring Tools and Significant Figures

An important factor in the accuracy and precision of measurements involves the precision of the measuring tool. In general, a precise measuring tool is one that can measure values in very small increments. For example, a standard ruler can measure length to the nearest millimeter, while a caliper can measure length to the nearest 0.01 millimeter. The caliper is a more precise measuring tool because it can measure extremely small differences in length. The more precise the measuring tool, the more precise and accurate the measurements can be.

When we express measured values, we can only list as many digits as we initially measured with our measuring tool. For example, if you use a standard ruler to measure the length of a stick, you may measure it to be 36.7 cm. You could not express this value as 36.71 cm because your measuring tool was not precise enough to measure a hundredth of a centimeter. It should be noted that the last digit in a measured value has been estimated in some way by the person performing the measurement. For example, the person measuring the length of a stick with a ruler notices that the stick length seems to be somewhere in between 36.6 cm and 36.7 cm, and he or she must estimate the value of the last digit. Using the

method of **significant figures**, the rule is that *the last digit written down in a measurement is the first digit with some uncertainty*. In order to determine the number of significant digits in a value, start with the first measured value at the left and count the number of digits through the last digit written on the right. For example, the measured value 36.7 cm has three digits, or significant figures. Significant figures indicate the precision of a measuring tool that was used to measure a value.

Zeros

Special consideration is given to zeros when counting significant figures. The zeros in 0.053 are not significant, because they are only placekeepers that locate the decimal point. There are two significant figures in 0.053. The zeros in 10.053 are not placekeepers but are significant—this number has five significant figures. The zeros in 1300 may or may not be significant depending on the style of writing numbers. They could mean the number is known to the last digit, or they could be placekeepers. So 1300 could have two, three, or four significant figures. (To avoid this ambiguity, write 1300 in scientific notation.) *Zeros are significant except when they serve only as placekeepers.*

Exercise:

Check Your Understanding

Problem:

Determine the number of significant figures in the following measurements:

- a. 0.0009
- b. 15,450.0
- c. 6×10^3
- d. 87.990
- e. 30.42

Solution:

- (a) 1; the zeros in this number are placekeepers that indicate the decimal point
- (b) 6; here, the zeros indicate that a measurement was made to the 0.1 decimal point, so the zeros are significant
- (c) 1; the value 10^3 signifies the decimal place, not the number of measured values
- (d) 5; the final zero indicates that a measurement was made to the 0.001 decimal point, so it is significant
- (e) 4; any zeros located in between significant figures in a number are also significant

Significant Figures in Calculations

When combining measurements with different degrees of accuracy and precision, *the number of significant digits in the final answer can be no greater than the number of significant digits in the least precise measured value*. There are two different rules, one for multiplication and division and the other for addition and subtraction, as discussed below.

1. For multiplication and division: *The result should have the same number of significant figures as the quantity having the least significant figures entering into the calculation.* For example, the area of a circle can be calculated from its radius using $A = \pi r^2$. Let us see how many significant figures the area has if the radius has only two—say, $r = 1.2$ m. Then,

Equation:

$$A = \pi r^2 = (3.1415927...) \times (1.2 \text{ m})^2 = 4.5238934 \text{ m}^2$$

is what you would get using a calculator that has an eight-digit output. But because the radius has only two significant figures, it limits the calculated

quantity to two significant figures or

Equation:

$$A=4.5 \text{ m}^2,$$

even though π is good to at least eight digits.

2. For addition and subtraction: *The answer can contain no more decimal places than the least precise measurement.* Suppose that you buy 7.56-kg of potatoes in a grocery store as measured with a scale with precision 0.01 kg. Then you drop off 6.052-kg of potatoes at your laboratory as measured by a scale with precision 0.001 kg. Finally, you go home and add 13.7 kg of potatoes as measured by a bathroom scale with precision 0.1 kg. How many kilograms of potatoes do you now have, and how many significant figures are appropriate in the answer? The mass is found by simple addition and subtraction:

Equation:

$$\begin{array}{r} 7.56 \text{ kg} \\ - 6.052 \text{ kg} \\ \hline +13.7 \text{ kg} \\ 15.208 \text{ kg} \end{array} = 15.2 \text{ kg}.$$

Next, we identify the least precise measurement: 13.7 kg. This measurement is expressed to the 0.1 decimal place, so our final answer must also be expressed to the 0.1 decimal place. Thus, the answer is rounded to the tenths place, giving us 15.2 kg.

Significant Figures in this Text

In this text, most numbers are assumed to have three significant figures. Furthermore, consistent numbers of significant figures are used in all worked examples. You will note that an answer given to three digits is based on input good to at least three digits, for example. If the input has fewer significant figures, the answer will also have fewer significant

figures. Care is also taken that the number of significant figures is reasonable for the situation posed. In some topics, particularly in optics, more accurate numbers are needed and more than three significant figures will be used. Finally, if a number is *exact*, such as the two in the formula for the circumference of a circle, $c = 2\pi r$, it does not affect the number of significant figures in a calculation.

Exercise:

Check Your Understanding

Problem:

Perform the following calculations and express your answer using the correct number of significant digits.

- (a) A woman has two bags weighing 13.5 pounds and one bag with a weight of 10.2 pounds. What is the total weight of the bags?
- (b) The force F on an object is equal to its mass m multiplied by its acceleration a . If a wagon with mass 55 kg accelerates at a rate of 0.0255 m/s^2 , what is the force on the wagon? (The unit of force is called the newton, and it is expressed with the symbol N.)

Solution:

- (a) 37.2 pounds; Because the number of bags is an exact value, it is not considered in the significant figures.
- (b) 1.4 N; Because the value 55 kg has only two significant figures, the final value must also contain two significant figures.

Note:

PhET Explorations: Estimation

Explore size estimation in one, two, and three dimensions! Multiple levels of difficulty allow for progressive skill improvement.

https://phet.colorado.edu/sims/estimation/estimation_en.html

Summary

- Accuracy of a measured value refers to how close a measurement is to the correct value. The uncertainty in a measurement is an estimate of the amount by which the measurement result may differ from this value.
- Precision of measured values refers to how close the agreement is between repeated measurements.
- The precision of a *measuring tool* is related to the size of its measurement increments. The smaller the measurement increment, the more precise the tool.
- Significant figures express the precision of a measuring tool.
- When multiplying or dividing measured values, the final answer can contain only as many significant figures as the least precise value.
- When adding or subtracting measured values, the final answer cannot contain more decimal places than the least precise value.

Conceptual Questions

Exercise:

Problem:

What is the relationship between the accuracy and uncertainty of a measurement?

Exercise:

Problem:

Prescriptions for vision correction are given in units called *diopters* (D). Determine the meaning of that unit. Obtain information (perhaps by calling an optometrist or performing an internet search) on the minimum uncertainty with which corrections in diopters are determined and the accuracy with which corrective lenses can be produced. Discuss the sources of uncertainties in both the prescription and accuracy in the manufacture of lenses.

Problems & Exercises

Express your answers to problems in this section to the correct number of significant figures and proper units.

Exercise:

Problem:

Suppose that your bathroom scale reads your mass as 65 kg with a 3% uncertainty. What is the uncertainty in your mass (in kilograms)?

Solution:

2 kg

Exercise:

Problem:

A good-quality measuring tape can be off by 0.50 cm over a distance of 20 m. What is its percent uncertainty?

Exercise:

Problem:

(a) A car speedometer has a 5.0% uncertainty. What is the range of possible speeds when it reads 90 km/h? (b) Convert this range to miles per hour. (1 km = 0.6214 mi)

Solution:

a. 85.5 to 94.5 km/h

b. 53.1 to 58.7 mi/h

Exercise:

Problem:

An infant's pulse rate is measured to be 130 ± 5 beats/min. What is the percent uncertainty in this measurement?

Exercise:**Problem:**

(a) Suppose that a person has an average heart rate of 72.0 beats/min. How many beats does he or she have in 2.0 y? (b) In 2.00 y? (c) In 2.000 y?

Solution:

(a) 7.6×10^7 beats

(b) 7.57×10^7 beats

(c) 7.57×10^7 beats

Exercise:**Problem:**

A can contains 375 mL of soda. How much is left after 308 mL is removed?

Exercise:**Problem:**

State how many significant figures are proper in the results of the following calculations: (a) $(106.7)(98.2)/(46.210)(1.01)$ (b) $(18.7)^2$ (c) $(1.60 \times 10^{-19})(3712)$.

Solution:

a. 3

b. 3

c. 3

Exercise:

Problem:

(a) How many significant figures are in the numbers 99 and 100? (b) If the uncertainty in each number is 1, what is the percent uncertainty in each? (c) Which is a more meaningful way to express the accuracy of these two numbers, significant figures or percent uncertainties?

Exercise:**Problem:**

(a) If your speedometer has an uncertainty of 2.0 km/h at a speed of 90 km/h, what is the percent uncertainty? (b) If it has the same percent uncertainty when it reads 60 km/h, what is the range of speeds you could be going?

Solution:

a) 2.2%

(b) 59 to 61 km/h

Exercise:**Problem:**

(a) A person's blood pressure is measured to be 120 ± 2 mm Hg. What is its percent uncertainty? (b) Assuming the same percent uncertainty, what is the uncertainty in a blood pressure measurement of 80 mm Hg?

Exercise:**Problem:**

A person measures his or her heart rate by counting the number of beats in 30 s. If 40 ± 1 beats are counted in 30.0 ± 0.5 s, what is the heart rate and its uncertainty in beats per minute?

Solution:

80 ± 3 beats/min

Exercise:

Problem: What is the area of a circle 3.102 cm in diameter?

Exercise:

Problem:

If a marathon runner averages 9.5 mi/h, how long does it take him or her to run a 26.22-mi marathon?

Solution:

2.8 h

Exercise:

Problem:

A marathon runner completes a 42.188-km course in 2 h, 30 min, and 12 s. There is an uncertainty of 25 m in the distance traveled and an uncertainty of 1 s in the elapsed time. (a) Calculate the percent uncertainty in the distance. (b) Calculate the uncertainty in the elapsed time. (c) What is the average speed in meters per second? (d) What is the uncertainty in the average speed?

Exercise:

Problem:

The sides of a small rectangular box are measured to be 1.80 ± 0.01 cm, 2.05 ± 0.02 cm, and 3.1 ± 0.1 cm long. Calculate its volume and uncertainty in cubic centimeters.

Solution:

11 ± 1 cm³

Exercise:

Problem:

When non-metric units were used in the United Kingdom, a unit of mass called the *pound-mass* (lbm) was employed, where $1 \text{ lbm} = 0.4539 \text{ kg}$. (a) If there is an uncertainty of 0.0001 kg in the pound-mass unit, what is its percent uncertainty? (b) Based on that percent uncertainty, what mass in pound-mass has an uncertainty of 1 kg when converted to kilograms?

Exercise:**Problem:**

The length and width of a rectangular room are measured to be $3.955 \pm 0.005 \text{ m}$ and $3.050 \pm 0.005 \text{ m}$. Calculate the area of the room and its uncertainty in square meters.

Solution:

$$12.06 \pm 0.04 \text{ m}^2$$

Exercise:**Problem:**

A car engine moves a piston with a circular cross section of $7.500 \pm 0.002 \text{ cm}$ diameter a distance of $3.250 \pm 0.001 \text{ cm}$ to compress the gas in the cylinder. (a) By what amount is the gas decreased in volume in cubic centimeters? (b) Find the uncertainty in this volume.

Glossary

accuracy

the degree to which a measured value agrees with correct value for that measurement

method of adding percents

the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation

percent uncertainty

the ratio of the uncertainty of a measurement to the measured value, expressed as a percentage

precision

the degree to which repeated measurements agree with each other

significant figures

express the precision of a measuring tool used to measure a value

uncertainty

a quantitative measure of how much your measured values deviate from a standard or expected value

Approximation

- Make reasonable approximations based on given data.

On many occasions, physicists, other scientists, and engineers need to make **approximations** or “guesstimates” for a particular quantity. What is the distance to a certain destination? What is the approximate density of a given item? About how large a current will there be in a circuit? Many approximate numbers are based on formulae in which the input quantities are known only to a limited accuracy. As you develop problem-solving skills (that can be applied to a variety of fields through a study of physics), you will also develop skills at approximating. You will develop these skills through thinking more quantitatively, and by being willing to take risks. As with any endeavor, experience helps, as well as familiarity with units. These approximations allow us to rule out certain scenarios or unrealistic numbers. Approximations also allow us to challenge others and guide us in our approaches to our scientific world. Let us do two examples to illustrate this concept.

Example:

Approximate the Height of a Building

Can you approximate the height of one of the buildings on your campus, or in your neighborhood? Let us make an approximation based upon the height of a person. In this example, we will calculate the height of a 39-story building.

Strategy

Think about the average height of an adult male. We can approximate the height of the building by scaling up from the height of a person.

Solution

Based on information in the example, we know there are 39 stories in the building. If we use the fact that the height of one story is approximately equal to about the length of two adult humans (each human is about 2-m tall), then we can estimate the total height of the building to be

Equation:

$$\frac{2 \text{ m}}{1 \text{ person}} \times \frac{2 \text{ person}}{1 \text{ story}} \times 39 \text{ stories} = 156 \text{ m.}$$

Discussion

You can use known quantities to determine an approximate measurement of unknown quantities. If your hand measures 10 cm across, how many hand lengths equal the width of your desk? What other measurements can you approximate besides length?

Example:

Approximating Vast Numbers: a Trillion Dollars



A bank stack contains one-hundred \$100 bills, and is worth \$10,000. How many bank stacks make up a trillion dollars? (credit: Andrew Magill)

The U.S. federal deficit in the 2008 fiscal year was a little greater than \$10 trillion. Most of us do not have any concept of how much even one trillion actually is. Suppose that you were given a trillion dollars in \$100 bills. If you made 100-bill stacks and used them to evenly cover a football field (between the end zones), make an approximation of how high the money pile would become. (We will use feet/inches rather than meters here)

because football fields are measured in yards.) One of your friends says 3 in., while another says 10 ft. What do you think?

Strategy

When you imagine the situation, you probably envision thousands of small stacks of 100 wrapped \$100 bills, such as you might see in movies or at a bank. Since this is an easy-to-approximate quantity, let us start there. We can find the volume of a stack of 100 bills, find out how many stacks make up one trillion dollars, and then set this volume equal to the area of the football field multiplied by the unknown height.

Solution

(1) Calculate the volume of a stack of 100 bills. The dimensions of a single bill are approximately 3 in. by 6 in. A stack of 100 of these is about 0.5 in. thick. So the total volume of a stack of 100 bills is:

Equation:

$$\begin{aligned}\text{volume of stack} &= \text{length} \times \text{width} \times \text{height}, \\ \text{volume of stack} &= 6 \text{ in.} \times 3 \text{ in.} \times 0.5 \text{ in.}, \\ \text{volume of stack} &= 9 \text{ in.}^3.\end{aligned}$$

(2) Calculate the number of stacks. Note that a trillion dollars is equal to $\$1 \times 10^{12}$, and a stack of one-hundred \$100 bills is equal to \$10,000, or $\$1 \times 10^4$. The number of stacks you will have is:

Equation:

$$\$1 \times 10^{12} (\text{a trillion dollars}) / \$1 \times 10^4 \text{ per stack} = 1 \times 10^8 \text{ stacks.}$$

(3) Calculate the area of a football field in square inches. The area of a football field is 100 yd \times 50 yd, which gives 5,000 yd². Because we are working in inches, we need to convert square yards to square inches:

Equation:

$$\begin{aligned}\text{Area} &= 5,000 \text{ yd}^2 \times \frac{3 \text{ ft}}{1 \text{ yd}} \times \frac{3 \text{ ft}}{1 \text{ yd}} \times \frac{12 \text{ in.}}{1 \text{ ft}} \times \frac{12 \text{ in.}}{1 \text{ ft}} = 6,480,000 \text{ in.}^2, \\ \text{Area} &\approx 6 \times 10^6 \text{ in.}^2.\end{aligned}$$

This conversion gives us $6 \times 10^6 \text{ in.}^2$ for the area of the field. (Note that we are using only one significant figure in these calculations.)

(4) Calculate the total volume of the bills. The volume of all the \$100-bill stacks is $9 \text{ in.}^3/\text{stack} \times 10^8 \text{ stacks} = 9 \times 10^8 \text{ in.}^3$.

(5) Calculate the height. To determine the height of the bills, use the equation:

Equation:

$$\text{volume of bills} = \text{area of field} \times \text{height of money:}$$

$$\text{Height of money} = \frac{\text{volume of bills}}{\text{area of field}},$$

$$\text{Height of money} = \frac{9 \times 10^8 \text{ in.}^3}{6 \times 10^6 \text{ in.}^2} = 1.33 \times 10^2 \text{ in.},$$

$$\text{Height of money} \approx 1 \times 10^2 \text{ in.} = 100 \text{ in.}$$

The height of the money will be about 100 in. high. Converting this value to feet gives

Equation:

$$100 \text{ in.} \times \frac{1 \text{ ft}}{12 \text{ in.}} = 8.33 \text{ ft} \approx 8 \text{ ft.}$$

Discussion

The final approximate value is much higher than the early estimate of 3 in., but the other early estimate of 10 ft (120 in.) was roughly correct. How did the approximation measure up to your first guess? What can this exercise tell you in terms of rough “guesstimates” versus carefully calculated approximations?

Exercise:

Check Your Understanding

Problem:

Using mental math and your understanding of fundamental units, approximate the area of a regulation basketball court. Describe the process you used to arrive at your final approximation.

Solution:

An average male is about two meters tall. It would take approximately 15 men laid out end to end to cover the length, and about 7 to cover the width. That gives an approximate area of 420 m^2 .

Summary

Scientists often approximate the values of quantities to perform calculations and analyze systems.

Problems & Exercises

Exercise:

Problem: How many heartbeats are there in a lifetime?

Solution:

Sample answer: 2×10^9 heartbeats

Exercise:

Problem:

A generation is about one-third of a lifetime. Approximately how many generations have passed since the year 0 AD?

Exercise:

Problem:

How many times longer than the mean life of an extremely unstable atomic nucleus is the lifetime of a human? (Hint: The lifetime of an unstable atomic nucleus is on the order of 10^{-22} s .)

Solution:

Sample answer: 2×10^{31} if an average human lifetime is taken to be about 70 years.

Exercise:

Problem:

Calculate the approximate number of atoms in a bacterium. Assume that the average mass of an atom in the bacterium is ten times the mass of a hydrogen atom. (Hint: The mass of a hydrogen atom is on the order of 10^{-27} kg and the mass of a bacterium is on the order of 10^{-15} kg.)



This color-enhanced photo shows *Salmonella typhimurium* (red) attacking human cells. These bacteria are commonly known for causing foodborne illness. Can you estimate the number of atoms in each bacterium? (credit: Rocky Mountain Laboratories, NIAID, NIH)

Exercise:

Problem:

Approximately how many atoms thick is a cell membrane, assuming all atoms there average about twice the size of a hydrogen atom?

Solution:

Sample answer: 50 atoms

Exercise:**Problem:**

(a) What fraction of Earth's diameter is the greatest ocean depth? (b) The greatest mountain height?

Exercise:**Problem:**

(a) Calculate the number of cells in a hummingbird assuming the mass of an average cell is ten times the mass of a bacterium. (b) Making the same assumption, how many cells are there in a human?

Solution:

Sample answers:

(a) 10^{12} cells/hummingbird

(b) 10^{16} cells/human

Exercise:**Problem:**

Assuming one nerve impulse must end before another can begin, what is the maximum firing rate of a nerve in impulses per second?

Glossary

approximation

an estimated value based on prior experience and reasoning

Introduction to One-Dimensional Kinematics

class="introduction"

The motion
of an
American
kestrel
through the
air can be
described by
the bird's
displacement
, speed,
velocity, and
acceleration.
When it flies
in a straight
line without
any change
in direction,
its motion is
said to be
one
dimensional.
(credit: Vince
Maidens,
Wikimedia
Commons)



Objects are in motion everywhere we look. Everything from a tennis game to a space-probe flyby of the planet Neptune involves motion. When you are resting, your heart moves blood through your veins. And even in inanimate objects, there is continuous motion in the vibrations of atoms and molecules. Questions about motion are interesting in and of themselves: *How long will it take for a space probe to get to Mars? Where will a football land if it is thrown at a certain angle?* But an understanding of motion is also key to understanding other concepts in physics. An understanding of acceleration, for example, is crucial to the study of force.

Our formal study of physics begins with **kinematics** which is defined as the *study of motion without considering its causes*. The word “kinematics” comes from a Greek term meaning motion and is related to other English words such as “cinema” (movies) and “kinesiology” (the study of human motion). In one-dimensional kinematics and [Two-Dimensional Kinematics](#) we will study only the *motion* of a football, for example, without worrying about what forces cause or change its motion. Such considerations come in other chapters. In this chapter, we examine the simplest type of motion—namely, motion along a straight line, or one-dimensional motion. In [Two-Dimensional Kinematics](#), we apply concepts developed here to study motion along curved paths (two- and three-dimensional motion); for example, that of a car rounding a curve.

Displacement

- Define position, displacement, distance, and distance traveled.
- Explain the relationship between position and displacement.
- Distinguish between displacement and distance traveled.
- Calculate displacement and distance given initial position, final position, and the path between the two.



These cyclists in Vietnam can be described by their position relative to buildings and a canal. Their motion can be described by their change in position, or displacement, in the frame of reference. (credit: Suzan Black, Fotopedia)

Position

In order to describe the motion of an object, you must first be able to describe its **position**—where it is at any particular time. More precisely, you need to specify its position relative to a convenient reference frame. Earth is often used as a reference frame, and we often describe the position of an object as it relates to stationary objects in that reference frame. For

example, a rocket launch would be described in terms of the position of the rocket with respect to the Earth as a whole, while a professor's position could be described in terms of where she is in relation to the nearby white board. (See [\[link\]](#).) In other cases, we use reference frames that are not stationary but are in motion relative to the Earth. To describe the position of a person in an airplane, for example, we use the airplane, not the Earth, as the reference frame. (See [\[link\]](#).)

Displacement

If an object moves relative to a reference frame (for example, if a professor moves to the right relative to a white board or a passenger moves toward the rear of an airplane), then the object's position changes. This change in position is known as **displacement**. The word “displacement” implies that an object has moved, or has been displaced.

Note:

Displacement

Displacement is the *change in position* of an object:

Equation:

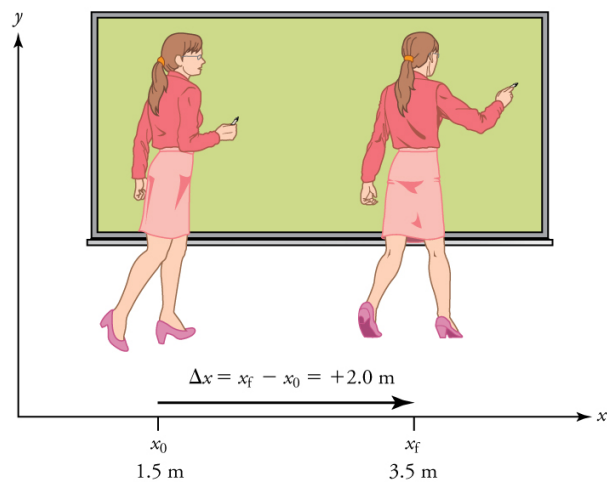
$$\Delta x = x_f - x_0,$$

where Δx is displacement, x_f is the final position, and x_0 is the initial position.

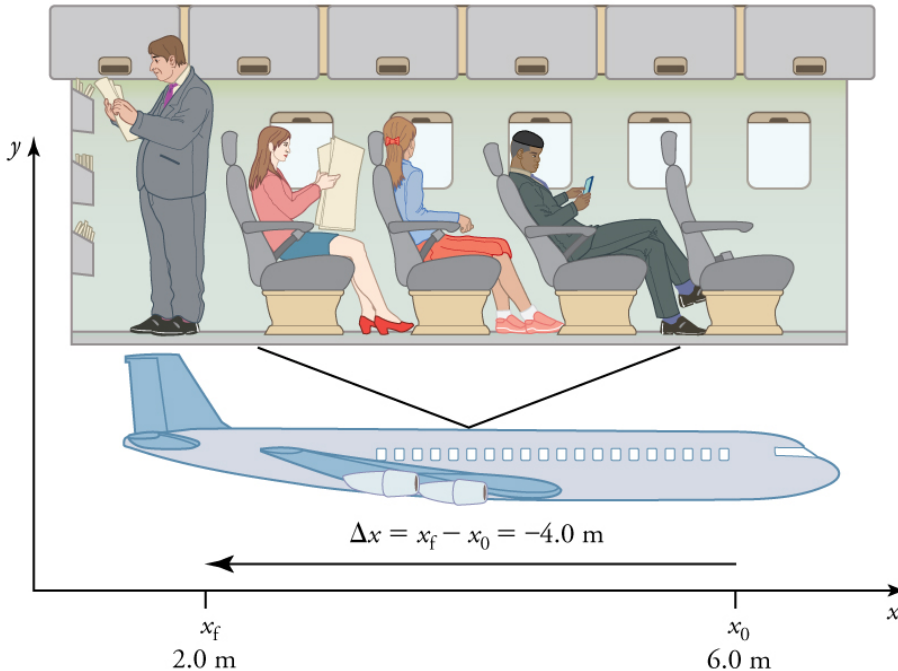
In this text the upper case Greek letter Δ (delta) always means “change in” whatever quantity follows it; thus, Δx means *change in position*. Always solve for displacement by subtracting initial position x_0 from final position x_f .

Note that the SI unit for displacement is the meter (m) (see [Physical Quantities and Units](#)), but sometimes kilometers, miles, feet, and other units of length are used. Keep in mind that when units other than the meter are

used in a problem, you may need to convert them into meters to complete the calculation.



A professor paces left and right while lecturing. Her position relative to Earth is given by x . The $+2.0 \text{ m}$ displacement of the professor relative to Earth is represented by an arrow pointing to the right.



A passenger moves from his seat to the back of the plane. His location relative to the airplane is given by x . The -4.0-m displacement of the passenger relative to the plane is represented by an arrow toward the rear of the plane. Notice that the arrow representing his displacement is twice as long as the arrow representing the displacement of the professor (he moves twice as far) in [\[link\]](#).

Note that displacement has a direction as well as a magnitude. The professor's displacement is 2.0 m to the right, and the airline passenger's displacement is 4.0 m toward the rear. In one-dimensional motion, direction can be specified with a plus or minus sign. When you begin a problem, you should select which direction is positive (usually that will be to the right or up, but you are free to select positive as being any direction). The professor's initial position is $x_0 = 1.5 \text{ m}$ and her final position is $x_f = 3.5 \text{ m}$. Thus her displacement is

Equation:

$$\Delta x = x_f - x_0 = 3.5 \text{ m} - 1.5 \text{ m} = +2.0 \text{ m}.$$

In this coordinate system, motion to the right is positive, whereas motion to the left is negative. Similarly, the airplane passenger's initial position is $x_0 = 6.0$ m and his final position is $x_f = 2.0$ m, so his displacement is

Equation:

$$\Delta x = x_f - x_0 = 2.0 \text{ m} - 6.0 \text{ m} = -4.0 \text{ m}.$$

His displacement is negative because his motion is toward the rear of the plane, or in the negative x direction in our coordinate system.

Distance

Although displacement is described in terms of direction, distance is not. **Distance** is defined to be *the magnitude or size of displacement between two positions*. Note that the distance between two positions is not the same as the distance traveled between them. **Distance traveled** is *the total length of the path traveled between two positions*. Distance has no direction and, thus, no sign. For example, the distance the professor walks is 2.0 m. The distance the airplane passenger walks is 4.0 m.

Note:

Misconception Alert: Distance Traveled vs. Magnitude of Displacement

It is important to note that the *distance traveled*, however, can be greater than the magnitude of the displacement (by magnitude, we mean just the size of the displacement without regard to its direction; that is, just a number with a unit). For example, the professor could pace back and forth many times, perhaps walking a distance of 150 m during a lecture, yet still end up only 2.0 m to the right of her starting point. In this case her displacement would be +2.0 m, the magnitude of her displacement would be 2.0 m, but the distance she traveled would be 150 m. In kinematics we nearly always deal with displacement and magnitude of displacement, and almost never with distance traveled. One way to think about this is to assume you marked the start of the motion and the end of the motion. The

displacement is simply the difference in the position of the two marks and is independent of the path taken in traveling between the two marks. The distance traveled, however, is the total length of the path taken between the two marks.

Exercise:

Check Your Understanding

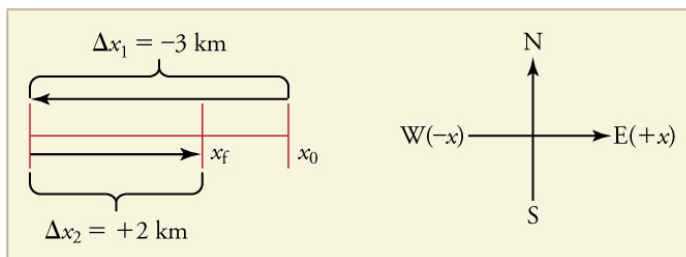
Problem:

A cyclist rides 3 km west and then turns around and rides 2 km east.

(a) What is her displacement? (b) What distance does she ride? (c)

What is the magnitude of her displacement?

Solution:



(a) The rider's displacement is $\Delta x = x_f - x_0 = -1 \text{ km}$. (The displacement is negative because we take east to be positive and west to be negative.)

(b) The distance traveled is $3 \text{ km} + 2 \text{ km} = 5 \text{ km}$.

(c) The magnitude of the displacement is 1 km .

Section Summary

- Kinematics is the study of motion without considering its causes. In this chapter, it is limited to motion along a straight line, called one-dimensional motion.
- Displacement is the change in position of an object.

- In symbols, displacement Δx is defined to be
Equation:

$$\Delta x = x_f - x_0,$$

where x_0 is the initial position and x_f is the final position. In this text, the Greek letter Δ (delta) always means “change in” whatever quantity follows it. The SI unit for displacement is the meter (m). Displacement has a direction as well as a magnitude.

- When you start a problem, assign which direction will be positive.
- Distance is the magnitude of displacement between two positions.
- Distance traveled is the total length of the path traveled between two positions.

Conceptual Questions

Exercise:

Problem:

Give an example in which there are clear distinctions among distance traveled, displacement, and magnitude of displacement. Specifically identify each quantity in your example.

Exercise:

Problem:

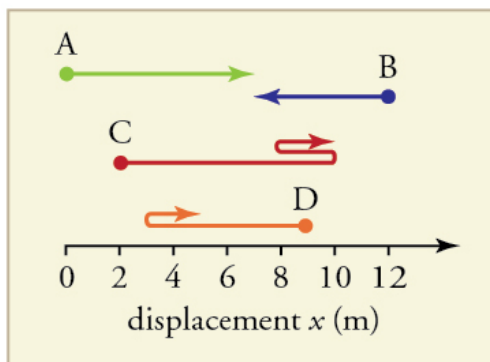
Under what circumstances does distance traveled equal magnitude of displacement? What is the only case in which magnitude of displacement and displacement are exactly the same?

Exercise:

Problem:

Bacteria move back and forth by using their flagella (structures that look like little tails). Speeds of up to $50 \mu\text{m/s}$ $50 \times 10^{-6} \text{ m/s}$ have been observed. The total distance traveled by a bacterium is large for its size, while its displacement is small. Why is this?

Problems & Exercises



Exercise:

Problem:

Find the following for path A in [\[link\]](#): (a) The distance traveled. (b) The magnitude of the displacement from start to finish. (c) The displacement from start to finish.

Solution:

(a) 7 m

(b) 7 m

(c) +7 m

Exercise:

Problem:

Find the following for path B in [\[link\]](#): (a) The distance traveled. (b) The magnitude of the displacement from start to finish. (c) The displacement from start to finish.

Exercise:

Problem:

Find the following for path C in [\[link\]](#): (a) The distance traveled. (b) The magnitude of the displacement from start to finish. (c) The displacement from start to finish.

Solution:

(a) 13 m

(b) 9 m

(c) +9 m

Exercise:**Problem:**

Find the following for path D in [\[link\]](#): (a) The distance traveled. (b) The magnitude of the displacement from start to finish. (c) The displacement from start to finish.

Glossary

kinematics

the study of motion without considering its causes

position

the location of an object at a particular time

displacement

the change in position of an object

distance

the magnitude of displacement between two positions

distance traveled

the total length of the path traveled between two positions

Vectors, Scalars, and Coordinate Systems

- Define and distinguish between scalar and vector quantities.
- Assign a coordinate system for a scenario involving one-dimensional motion.



The motion of this Eclipse Concept jet can be described in terms of the distance it has traveled (a scalar quantity) or its displacement in a specific direction (a vector quantity). In order to specify the direction of motion, its displacement must be described based on a coordinate system. In this case, it may be convenient to choose motion toward the left as positive motion (it is the forward direction for the plane), although in many cases, the x -coordinate runs from left to right, with motion to the right as positive and motion to the left as negative. (credit: Armchair Aviator, Flickr)

What is the difference between distance and displacement? Whereas displacement is defined by both direction and magnitude, distance is defined only by magnitude. Displacement is an example of a vector quantity. Distance is an example of a scalar quantity. A **vector** is any quantity with both *magnitude and direction*. Other examples of vectors include a velocity of 90 km/h east and a force of 500 newtons straight down.

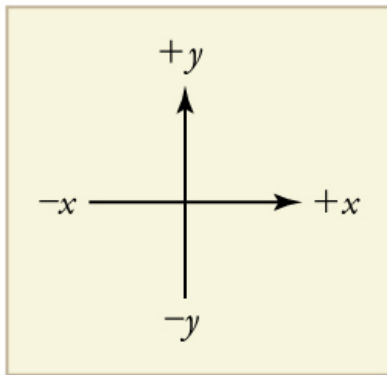
The direction of a vector in one-dimensional motion is given simply by a plus (+) or minus (−) sign. Vectors are represented graphically by arrows. An arrow used to represent a vector has a length proportional to the vector's magnitude (e.g., the larger the magnitude, the longer the length of the vector) and points in the same direction as the vector.

Some physical quantities, like distance, either have no direction or none is specified. A **scalar** is any quantity that has a magnitude, but no direction. For example, a 20°C temperature, the 250 kilocalories (250 Calories) of energy in a candy bar, a 90 km/h speed limit, a person's 1.8 m height, and a distance of 2.0 m are all scalars—quantities with no specified direction. Note, however, that a scalar can be negative, such as a −20°C temperature. In this case, the minus sign indicates a point on a scale rather than a direction. Scalars are never represented by arrows.

Coordinate Systems for One-Dimensional Motion

In order to describe the direction of a vector quantity, you must designate a coordinate system within the reference frame. For one-dimensional motion, this is a simple coordinate system consisting of a one-dimensional coordinate line. In general, when describing horizontal motion, motion to the right is usually considered positive, and motion to the left is considered negative. With vertical motion, motion up is usually positive and motion down is negative. In some cases, however, as with the jet in [\[link\]](#), it can be more convenient to switch the positive and negative directions. For example, if you are analyzing the motion of falling objects, it can be useful to define downwards as the positive direction. If people in a race are

running to the left, it is useful to define left as the positive direction. It does not matter as long as the system is clear and consistent. Once you assign a positive direction and start solving a problem, you cannot change it.



It is usually convenient to consider motion upward or to the right as positive (+) and motion downward or to the left as negative (−).

Exercise:

Check Your Understanding

Problem:

A person's speed can stay the same as he or she rounds a corner and changes direction. Given this information, is speed a scalar or a vector quantity? Explain.

Solution:

Speed is a scalar quantity. It does not change at all with direction changes; therefore, it has magnitude only. If it were a vector quantity, it would change as direction changes (even if its magnitude remained constant).

Section Summary

- A vector is any quantity that has magnitude and direction.
- A scalar is any quantity that has magnitude but no direction.
- Displacement and velocity are vectors, whereas distance and speed are scalars.
- In one-dimensional motion, direction is specified by a plus or minus sign to signify left or right, up or down, and the like.

Conceptual Questions

Exercise:

Problem:

A student writes, “A bird that is diving for prey has a speed of -10 m/s .” What is wrong with the student’s statement? What has the student actually described? Explain.

Exercise:

Problem: What is the speed of the bird in [\[link\]](#)?

Exercise:

Problem:

Acceleration is the change in velocity over time. Given this information, is acceleration a vector or a scalar quantity? Explain.

Exercise:

Problem:

A weather forecast states that the temperature is predicted to be -5°C the following day. Is this temperature a vector or a scalar quantity? Explain.

Glossary

scalar

a quantity that is described by magnitude, but not direction

vector

a quantity that is described by both magnitude and direction

Time, Velocity, and Speed

- Explain the relationships between instantaneous velocity, average velocity, instantaneous speed, average speed, displacement, and time.
- Calculate velocity and speed given initial position, initial time, final position, and final time.
- Derive a graph of velocity vs. time given a graph of position vs. time.
- Interpret a graph of velocity vs. time.



The motion of these racing snails can be described by their speeds and their velocities.
(credit: tobiasflickr, Flickr)

There is more to motion than distance and displacement. Questions such as, “How long does a foot race take?” and “What was the runner’s speed?” cannot be answered without an understanding of other concepts. In this section we add definitions of time, velocity, and speed to expand our description of motion.

Time

As discussed in [Physical Quantities and Units](#), the most fundamental physical quantities are defined by how they are measured. This is the case with time. Every measurement of time involves measuring a change in

some physical quantity. It may be a number on a digital clock, a heartbeat, or the position of the Sun in the sky. In physics, the definition of time is simple—**time** is *change*, or the interval over which change occurs. It is impossible to know that time has passed unless something changes.

The amount of time or change is calibrated by comparison with a standard. The SI unit for time is the second, abbreviated s. We might, for example, observe that a certain pendulum makes one full swing every 0.75 s. We could then use the pendulum to measure time by counting its swings or, of course, by connecting the pendulum to a clock mechanism that registers time on a dial. This allows us to not only measure the amount of time, but also to determine a sequence of events.

How does time relate to motion? We are usually interested in elapsed time for a particular motion, such as how long it takes an airplane passenger to get from his seat to the back of the plane. To find elapsed time, we note the time at the beginning and end of the motion and subtract the two. For example, a lecture may start at 11:00 A.M. and end at 11:50 A.M., so that the elapsed time would be 50 min. **Elapsed time** Δt is the difference between the ending time and beginning time,

Equation:

$$\Delta t = t_f - t_0,$$

where Δt is the change in time or elapsed time, t_f is the time at the end of the motion, and t_0 is the time at the beginning of the motion. (As usual, the delta symbol, Δ , means the change in the quantity that follows it.)

Life is simpler if the beginning time t_0 is taken to be zero, as when we use a stopwatch. If we were using a stopwatch, it would simply read zero at the start of the lecture and 50 min at the end. If $t_0 = 0$, then $\Delta t = t_f \equiv t$.

In this text, for simplicity's sake,

- motion starts at time equal to zero ($t_0 = 0$)
- the symbol t is used for elapsed time unless otherwise specified ($\Delta t = t_f \equiv t$)

Velocity

Your notion of velocity is probably the same as its scientific definition. You know that if you have a large displacement in a small amount of time you have a large velocity, and that velocity has units of distance divided by time, such as miles per hour or kilometers per hour.

Note:

Average Velocity

Average velocity is *displacement (change in position) divided by the time of travel*,

Equation:

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_f - x_0}{t_f - t_0},$$

where \bar{v} is the *average* (indicated by the bar over the v) velocity, Δx is the change in position (or displacement), and x_f and x_0 are the final and beginning positions at times t_f and t_0 , respectively. If the starting time t_0 is taken to be zero, then the average velocity is simply

Equation:

$$\bar{v} = \frac{\Delta x}{t}.$$

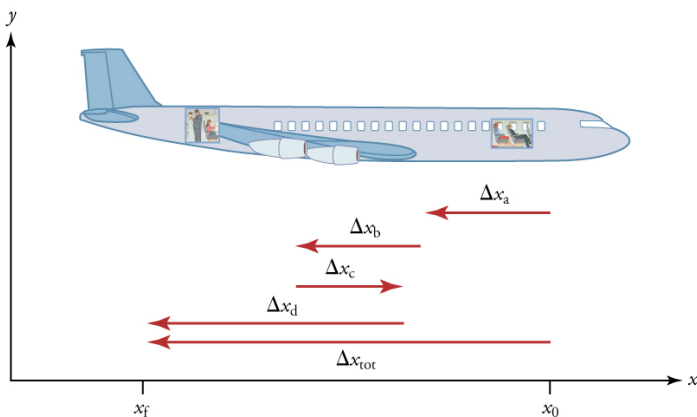
Notice that this definition indicates that *velocity is a vector because displacement is a vector*. It has both magnitude and direction. The SI unit for velocity is meters per second or m/s, but many other units, such as km/h, mi/h (also written as mph), and cm/s, are in common use. Suppose, for example, an airplane passenger took 5 seconds to move -4 m (the negative sign indicates that displacement is toward the back of the plane). His average velocity would be

Equation:

$$\bar{v} = \frac{\Delta x}{t} = \frac{-4 \text{ m}}{5 \text{ s}} = -0.8 \text{ m/s}.$$

The minus sign indicates the average velocity is also toward the rear of the plane.

The average velocity of an object does not tell us anything about what happens to it between the starting point and ending point, however. For example, we cannot tell from average velocity whether the airplane passenger stops momentarily or backs up before he goes to the back of the plane. To get more details, we must consider smaller segments of the trip over smaller time intervals.



A more detailed record of an airplane passenger heading toward the back of the plane, showing smaller segments of his trip.

The smaller the time intervals considered in a motion, the more detailed the information. When we carry this process to its logical conclusion, we are left with an infinitesimally small interval. Over such an interval, the average velocity becomes the *instantaneous velocity* or the *velocity at a specific instant*. A car's speedometer, for example, shows the magnitude (but not the

direction) of the instantaneous velocity of the car. (Police give tickets based on instantaneous velocity, but when calculating how long it will take to get from one place to another on a road trip, you need to use average velocity.)

Instantaneous velocity v is the average velocity at a specific instant in time (or over an infinitesimally small time interval).

Mathematically, finding instantaneous velocity, v , at a precise instant t can involve taking a limit, a calculus operation beyond the scope of this text.

However, under many circumstances, we can find precise values for instantaneous velocity without calculus.

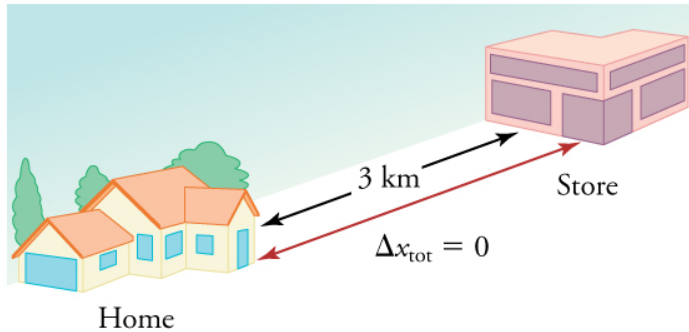
Speed

In everyday language, most people use the terms “speed” and “velocity” interchangeably. In physics, however, they do not have the same meaning and they are distinct concepts. One major difference is that speed has no direction. Thus *speed is a scalar*. Just as we need to distinguish between instantaneous velocity and average velocity, we also need to distinguish between instantaneous speed and average speed.

Instantaneous speed is the magnitude of instantaneous velocity. For example, suppose the airplane passenger at one instant had an instantaneous velocity of -3.0 m/s (the minus meaning toward the rear of the plane). At that same time his instantaneous speed was 3.0 m/s. Or suppose that at one time during a shopping trip your instantaneous velocity is 40 km/h due north. Your instantaneous speed at that instant would be 40 km/h—the same magnitude but without a direction. Average speed, however, is very different from average velocity. **Average speed** is the distance traveled divided by elapsed time.

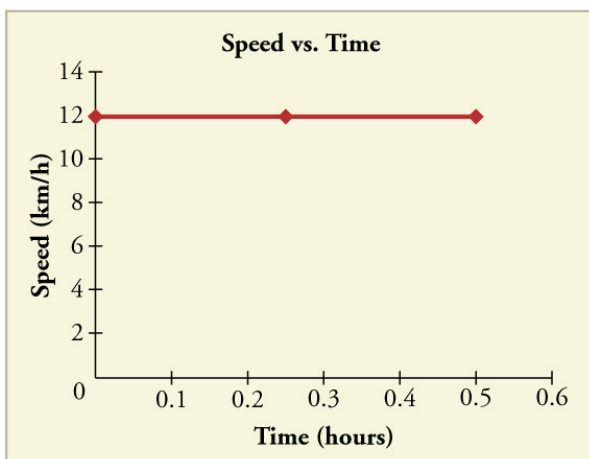
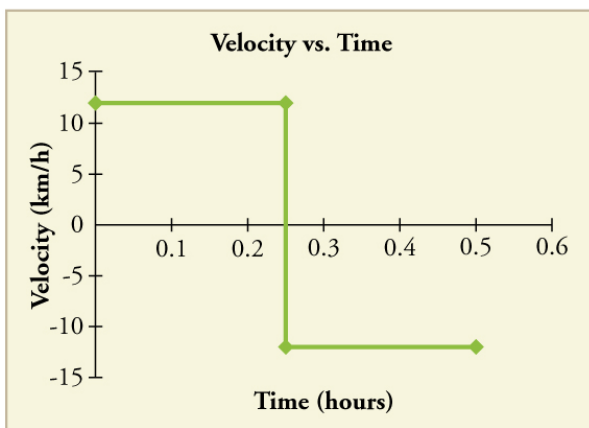
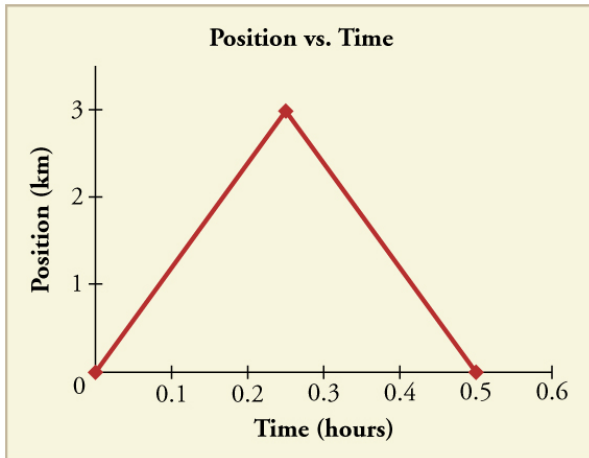
We have noted that distance traveled can be greater than displacement. So average speed can be greater than average velocity, which is displacement divided by time. For example, if you drive to a store and return home in half an hour, and your car’s odometer shows the total distance traveled was 6 km, then your average speed was 12 km/h. Your average velocity, however, was zero, because your displacement for the round trip is zero.

(Displacement is change in position and, thus, is zero for a round trip.) Thus average speed is *not* simply the magnitude of average velocity.



During a 30-minute round trip to the store, the total distance traveled is 6 km. The average speed is 12 km/h. The displacement for the round trip is zero, since there was no net change in position. Thus the average velocity is zero.

Another way of visualizing the motion of an object is to use a graph. A plot of position or of velocity as a function of time can be very useful. For example, for this trip to the store, the position, velocity, and speed-vs.-time graphs are displayed in [\[link\]](#). (Note that these graphs depict a very simplified **model** of the trip. We are assuming that speed is constant during the trip, which is unrealistic given that we'll probably stop at the store. But for simplicity's sake, we will model it with no stops or changes in speed. We are also assuming that the route between the store and the house is a perfectly straight line.)



Position vs. time, velocity vs. time, and speed vs. time on a trip. Note that the velocity for the return trip is negative.

Note:**Making Connections: Take-Home Investigation—Getting a Sense of Speed**

If you have spent much time driving, you probably have a good sense of speeds between about 10 and 70 miles per hour. But what are these in meters per second? What do we mean when we say that something is moving at 10 m/s? To get a better sense of what these values really mean, do some observations and calculations on your own:

- calculate typical car speeds in meters per second
- estimate jogging and walking speed by timing yourself; convert the measurements into both m/s and mi/h
- determine the speed of an ant, snail, or falling leaf

Exercise:**Check Your Understanding****Problem:**

A commuter train travels from Baltimore to Washington, DC, and back in 1 hour and 45 minutes. The distance between the two stations is approximately 40 miles. What is (a) the average velocity of the train, and (b) the average speed of the train in m/s?

Solution:

(a) The average velocity of the train is zero because $x_f = x_0$; the train ends up at the same place it starts.

(b) The average speed of the train is calculated below. Note that the train travels 40 miles one way and 40 miles back, for a total distance of 80 miles.

Equation:

$$\frac{\text{distance}}{\text{time}} = \frac{80 \text{ miles}}{105 \text{ minutes}}$$

Equation:

$$\frac{80 \text{ miles}}{105 \text{ minutes}} \times \frac{5280 \text{ feet}}{1 \text{ mile}} \times \frac{1 \text{ meter}}{3.28 \text{ feet}} \times \frac{1 \text{ minute}}{60 \text{ seconds}} = 20 \text{ m/s}$$

Section Summary

- Time is measured in terms of change, and its SI unit is the second (s). Elapsed time for an event is

Equation:

$$\Delta t = t_f - t_0,$$

where t_f is the final time and t_0 is the initial time. The initial time is often taken to be zero, as if measured with a stopwatch; the elapsed time is then just t .

- Average velocity \bar{v} is defined as displacement divided by the travel time. In symbols, average velocity is

Equation:

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_f - x_0}{t_f - t_0}.$$

- The SI unit for velocity is m/s.
- Velocity is a vector and thus has a direction.
- Instantaneous velocity v is the velocity at a specific instant or the average velocity for an infinitesimal interval.
- Instantaneous speed is the magnitude of the instantaneous velocity.
- Instantaneous speed is a scalar quantity, as it has no direction specified.
- Average speed is the total distance traveled divided by the elapsed time. (Average speed is *not* the magnitude of the average velocity.) Speed is a scalar quantity; it has no direction associated with it.

Conceptual Questions

Exercise:**Problem:**

Give an example (but not one from the text) of a device used to measure time and identify what change in that device indicates a change in time.

Exercise:**Problem:**

There is a distinction between average speed and the magnitude of average velocity. Give an example that illustrates the difference between these two quantities.

Exercise:**Problem:**

Does a car's odometer measure position or displacement? Does its speedometer measure speed or velocity?

Exercise:**Problem:**

If you divide the total distance traveled on a car trip (as determined by the odometer) by the time for the trip, are you calculating the average speed or the magnitude of the average velocity? Under what circumstances are these two quantities the same?

Exercise:**Problem:**

How are instantaneous velocity and instantaneous speed related to one another? How do they differ?

Problems & Exercises**Exercise:**

Problem:

(a) Calculate Earth's average speed relative to the Sun. (b) What is its average velocity over a period of one year?

Solution:

(a) $3.0 \times 10^4 \text{ m/s}$

(b) 0 m/s

Exercise:**Problem:**

A helicopter blade spins at exactly 100 revolutions per minute. Its tip is 5.00 m from the center of rotation. (a) Calculate the average speed of the blade tip in the helicopter's frame of reference. (b) What is its average velocity over one revolution?

Exercise:**Problem:**

The North American and European continents are moving apart at a rate of about 3 cm/y. At this rate how long will it take them to drift 500 km farther apart than they are at present?

Solution:

2×10^7 years

Exercise:

Problem:

Land west of the San Andreas fault in southern California is moving at an average velocity of about 6 cm/y northwest relative to land east of the fault. Los Angeles is west of the fault and may thus someday be at the same latitude as San Francisco, which is east of the fault. How far in the future will this occur if the displacement to be made is 590 km northwest, assuming the motion remains constant?

Exercise:**Problem:**

On May 26, 1934, a streamlined, stainless steel diesel train called the Zephyr set the world's nonstop long-distance speed record for trains. Its run from Denver to Chicago took 13 hours, 4 minutes, 58 seconds, and was witnessed by more than a million people along the route. The total distance traveled was 1633.8 km. What was its average speed in km/h and m/s?

Solution:

$$34.689 \text{ m/s} = 124.88 \text{ km/h}$$

Exercise:**Problem:**

Tidal friction is slowing the rotation of the Earth. As a result, the orbit of the Moon is increasing in radius at a rate of approximately 4 cm/year. Assuming this to be a constant rate, how many years will pass before the radius of the Moon's orbit increases by $3.84 \times 10^6 \text{ m}$ (1%)?

Exercise:

Problem:

A student drove to the university from her home and noted that the odometer reading of her car increased by 12.0 km. The trip took 18.0 min. (a) What was her average speed? (b) If the straight-line distance from her home to the university is 10.3 km in a direction 25.0° south of east, what was her average velocity? (c) If she returned home by the same path 7 h 30 min after she left, what were her average speed and velocity for the entire trip?

Solution:

(a) 40.0 km/h

(b) 34.3 km/h, 25° S of E.

(c) average speed = 3.20 km/h, $\bar{v} = 0$.

Exercise:**Problem:**

The speed of propagation of the action potential (an electrical signal) in a nerve cell depends (inversely) on the diameter of the axon (nerve fiber). If the nerve cell connecting the spinal cord to your feet is 1.1 m long, and the nerve impulse speed is 18 m/s, how long does it take for the nerve signal to travel this distance?

Exercise:

Problem:

Conversations with astronauts on the lunar surface were characterized by a kind of echo in which the earthbound person's voice was so loud in the astronaut's space helmet that it was picked up by the astronaut's microphone and transmitted back to Earth. It is reasonable to assume that the echo time equals the time necessary for the radio wave to travel from the Earth to the Moon and back (that is, neglecting any time delays in the electronic equipment). Calculate the distance from Earth to the Moon given that the echo time was 2.56 s and that radio waves travel at the speed of light (3.00×10^8 m/s).

Solution:

384,000 km

Exercise:**Problem:**

A football quarterback runs 15.0 m straight down the playing field in 2.50 s. He is then hit and pushed 3.00 m straight backward in 1.75 s. He breaks the tackle and runs straight forward another 21.0 m in 5.20 s. Calculate his average velocity (a) for each of the three intervals and (b) for the entire motion.

Exercise:**Problem:**

The planetary model of the atom pictures electrons orbiting the atomic nucleus much as planets orbit the Sun. In this model you can view hydrogen, the simplest atom, as having a single electron in a circular orbit 1.06×10^{-10} m in diameter. (a) If the average speed of the electron in this orbit is known to be 2.20×10^6 m/s, calculate the number of revolutions per second it makes about the nucleus. (b) What is the electron's average velocity?

Solution:

(a) $6.61 \times 10^{15} \text{ rev/s}$

(b) 0 m/s

Glossary

average speed

distance traveled divided by time during which motion occurs

average velocity

displacement divided by time over which displacement occurs

instantaneous velocity

velocity at a specific instant, or the average velocity over an infinitesimal time interval

instantaneous speed

magnitude of the instantaneous velocity

time

change, or the interval over which change occurs

model

simplified description that contains only those elements necessary to describe the physics of a physical situation

elapsed time

the difference between the ending time and beginning time

Acceleration

- Define and distinguish between instantaneous acceleration, average acceleration, and deceleration.
- Calculate acceleration given initial time, initial velocity, final time, and final velocity.



A plane decelerates, or slows down, as it comes in for landing in St. Maarten. Its acceleration is opposite in direction to its velocity. (credit: Steve Conry, Flickr)

In everyday conversation, to accelerate means to speed up. The accelerator in a car can in fact cause it to speed up. The greater the **acceleration**, the greater the change in velocity over a given time. The formal definition of acceleration is consistent with these notions, but more inclusive.

Note:

Average Acceleration

Average Acceleration is *the rate at which velocity changes*,

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0},$$

where \bar{a} is average acceleration, v is velocity, and t is time. (The bar over the a means *average* acceleration.)

Because acceleration is velocity in m/s divided by time in s, the SI units for acceleration are m/s^2 , meters per second squared or meters per second per second, which literally means by how many meters per second the velocity changes every second.

Recall that velocity is a vector—it has both magnitude and direction. This means that a change in velocity can be a change in magnitude (or speed), but it can also be a change in *direction*. For example, if a car turns a corner at constant speed, it is accelerating because its direction is changing. The quicker you turn, the greater the acceleration. So there is an acceleration when velocity changes either in magnitude (an increase or decrease in speed) or in direction, or both.

Note:

Acceleration as a Vector

Acceleration is a vector in the same direction as the *change* in velocity, Δv . Since velocity is a vector, it can change either in magnitude or in direction. Acceleration is therefore a change in either speed or direction, or both.

Keep in mind that although acceleration is in the direction of the *change* in velocity, it is not always in the direction of *motion*. When an object slows down, its acceleration is opposite to the direction of its motion. This is known as **deceleration**.

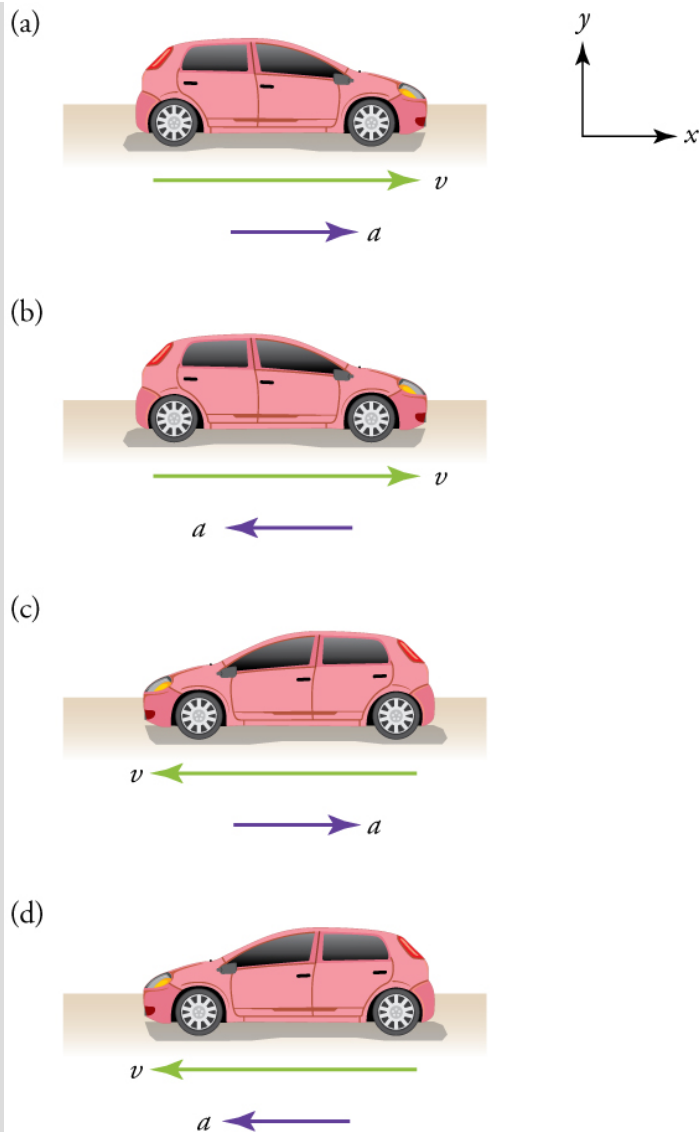


A subway train in Sao Paulo, Brazil, decelerates as it comes into a station. It is accelerating in a direction opposite to its direction of motion. (credit: Yusuke Kawasaki, Flickr)

Note:

Misconception Alert: Deceleration vs. Negative Acceleration

Deceleration always refers to acceleration in the direction opposite to the direction of the velocity. Deceleration always reduces speed. Negative acceleration, however, is acceleration *in the negative direction in the chosen coordinate system*. Negative acceleration may or may not be deceleration, and deceleration may or may not be considered negative acceleration. For example, consider [\[link\]](#).



(a) This car is speeding up as it moves toward the right. It therefore has positive acceleration in our coordinate system. (b) This car is slowing down as it moves toward the right. Therefore, it has negative acceleration in our coordinate system, because its acceleration is toward the left. The car is also decelerating: the direction of its acceleration is opposite to its direction of motion. (c) This car is moving

toward the left, but slowing down over time. Therefore, its acceleration is positive in our coordinate system because it is toward the right.

However, the car is decelerating because its acceleration is opposite to its motion. (d) This car is speeding up as it moves toward the left. It has negative acceleration because it is accelerating toward the left. However, because its acceleration is in the same direction as its motion, it is speeding up (*not* decelerating).

Example:

Calculating Acceleration: A Racehorse Leaves the Gate

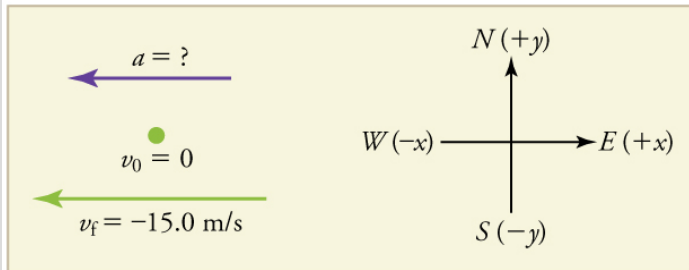
A racehorse coming out of the gate accelerates from rest to a velocity of 15.0 m/s due west in 1.80 s . What is its average acceleration?



(credit: Jon Sullivan, PD
Photo.org)

Strategy

First we draw a sketch and assign a coordinate system to the problem. This is a simple problem, but it always helps to visualize it. Notice that we assign east as positive and west as negative. Thus, in this case, we have negative velocity.



We can solve this problem by identifying Δv and Δt from the given information and then calculating the average acceleration directly from the equation $\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0}$.

Solution

1. Identify the knowns. $v_0 = 0$, $v_f = -15.0 \text{ m/s}$ (the negative sign indicates direction toward the west), $\Delta t = 1.80 \text{ s}$.
2. Find the change in velocity. Since the horse is going from zero to -15.0 m/s , its change in velocity equals its final velocity:

$$\Delta v = v_f = -15.0 \text{ m/s}.$$

3. Plug in the known values (Δv and Δt) and solve for the unknown \bar{a} .

Equation:

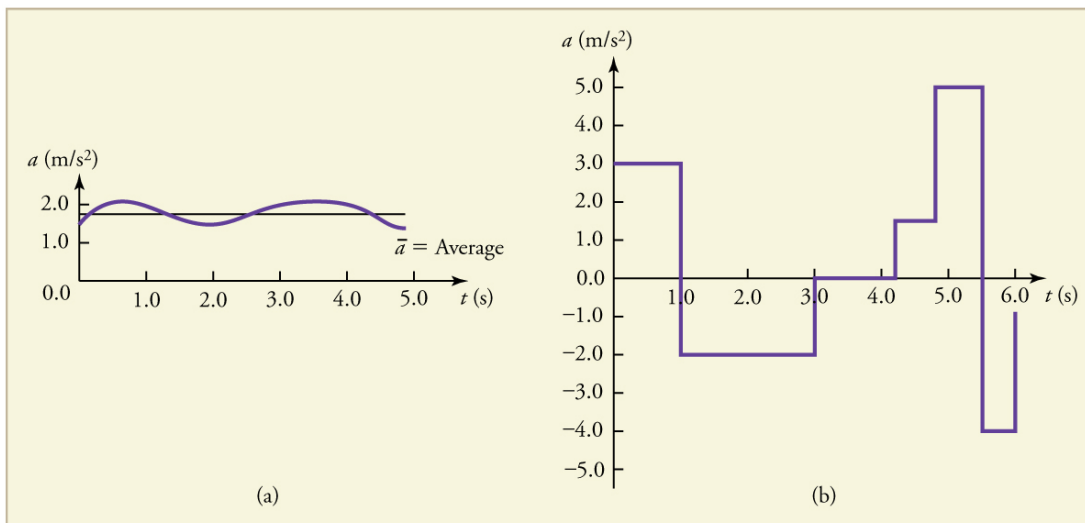
$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{-15.0 \text{ m/s}}{1.80 \text{ s}} = -8.33 \text{ m/s}^2.$$

Discussion

The negative sign for acceleration indicates that acceleration is toward the west. An acceleration of 8.33 m/s^2 due west means that the horse increases its velocity by 8.33 m/s due west each second, that is, 8.33 meters per second per second, which we write as 8.33 m/s^2 . This is truly an average acceleration, because the ride is not smooth. We shall see later that an acceleration of this magnitude would require the rider to hang on with a force nearly equal to his weight.

Instantaneous Acceleration

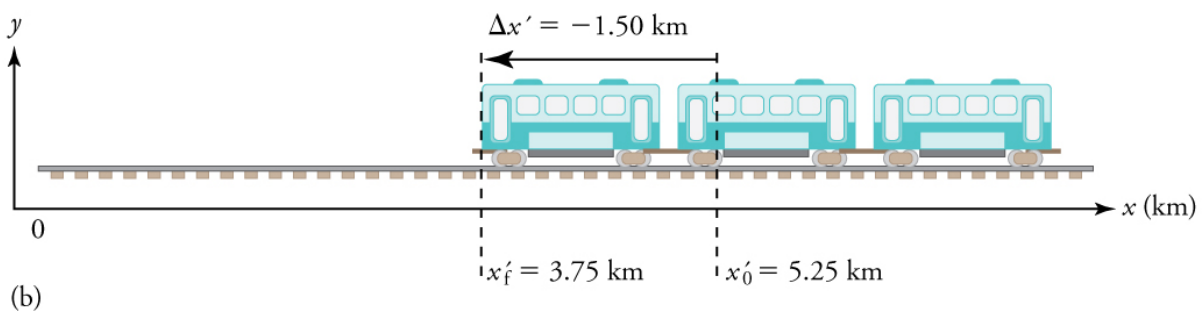
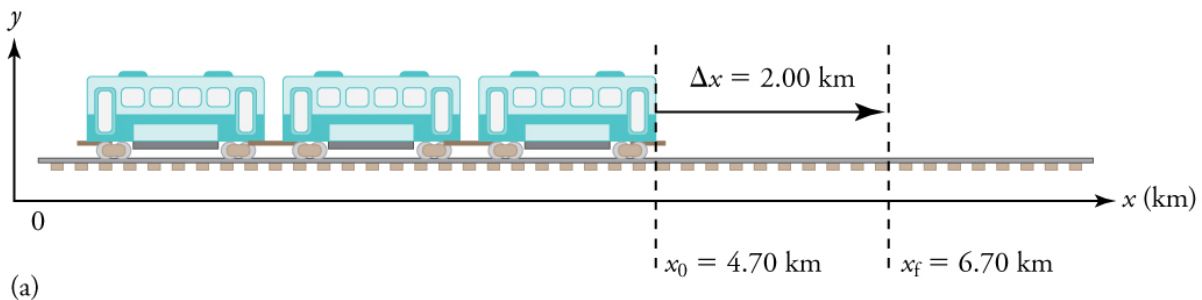
Instantaneous acceleration a , or the *acceleration at a specific instant in time*, is obtained by the same process as discussed for instantaneous velocity in [Time, Velocity, and Speed](#)—that is, by considering an infinitesimally small interval of time. How do we find instantaneous acceleration using only algebra? The answer is that we choose an average acceleration that is representative of the motion. [\[link\]](#) shows graphs of instantaneous acceleration versus time for two very different motions. In [\[link\]](#)(a), the acceleration varies slightly and the average over the entire interval is nearly the same as the instantaneous acceleration at any time. In [\[link\]](#)(b), the acceleration varies drastically over time. In such situations it is best to consider smaller time intervals and choose an average acceleration for each. For example, we could consider motion over the time intervals from 0 to 1.0 s and from 1.0 to 3.0 s as separate motions with accelerations of $+3.0 \text{ m/s}^2$ and -2.0 m/s^2 , respectively.



Graphs of instantaneous acceleration versus time for two different one-dimensional motions. (a) Here acceleration varies only slightly and is always in the same direction, since it is positive. The average over the interval is nearly the same as the

acceleration at any given time. (b) Here the acceleration varies greatly, perhaps representing a package on a post office conveyor belt that is accelerated forward and backward as it bumps along. It is necessary to consider small time intervals (such as from 0 to 1.0 s) with constant or nearly constant acceleration in such a situation.

The next several examples consider the motion of the subway train shown in [\[link\]](#). In (a) the shuttle moves to the right, and in (b) it moves to the left. The examples are designed to further illustrate aspects of motion and to illustrate some of the reasoning that goes into solving problems.



One-dimensional motion of a subway train considered in [\[link\]](#), [\[link\]](#), [\[link\]](#), [\[link\]](#), [\[link\]](#), and [\[link\]](#). Here we have chosen the x -axis so that $+$ means to the right and $-$ means to the left for displacements, velocities, and accelerations. (a) The subway train moves to the right from x_0 to x_f . Its displacement Δx is $+2.0$ km. (b) The train moves to the left from x'_0 to x'_f . Its displacement $\Delta x'$ is

–1.5 km. (Note that the prime symbol (') is used simply to distinguish between displacement in the two different situations. The distances of travel and the size of the cars are on different scales to fit everything into the diagram.)

Example:

Calculating Displacement: A Subway Train

What are the magnitude and sign of displacements for the motions of the subway train shown in parts (a) and (b) of [\[link\]](#)?

Strategy

A drawing with a coordinate system is already provided, so we don't need to make a sketch, but we should analyze it to make sure we understand what it is showing. Pay particular attention to the coordinate system. To find displacement, we use the equation $\Delta x = x_f - x_0$. This is straightforward since the initial and final positions are given.

Solution

1. Identify the knowns. In the figure we see that $x_f = 6.70$ km and $x_0 = 4.70$ km for part (a), and $x'_f = 3.75$ km and $x'_0 = 5.25$ km for part (b).
2. Solve for displacement in part (a).

Equation:

$$\Delta x = x_f - x_0 = 6.70 \text{ km} - 4.70 \text{ km} = +2.00 \text{ km}$$

3. Solve for displacement in part (b).

Equation:

$$\Delta x' = x'_f - x'_0 = 3.75 \text{ km} - 5.25 \text{ km} = -1.50 \text{ km}$$

Discussion

The direction of the motion in (a) is to the right and therefore its displacement has a positive sign, whereas motion in (b) is to the left and thus has a negative sign.

Example:**Comparing Distance Traveled with Displacement: A Subway Train**

What are the distances traveled for the motions shown in parts (a) and (b) of the subway train in [\[link\]](#)?

Strategy

To answer this question, think about the definitions of distance and distance traveled, and how they are related to displacement. Distance between two positions is defined to be the magnitude of displacement, which was found in [\[link\]](#). Distance traveled is the total length of the path traveled between the two positions. (See [Displacement](#).) In the case of the subway train shown in [\[link\]](#), the distance traveled is the same as the distance between the initial and final positions of the train.

Solution

1. The displacement for part (a) was $+2.00$ km. Therefore, the distance between the initial and final positions was 2.00 km, and the distance traveled was 2.00 km.
2. The displacement for part (b) was -1.5 km. Therefore, the distance between the initial and final positions was 1.50 km, and the distance traveled was 1.50 km.

Discussion

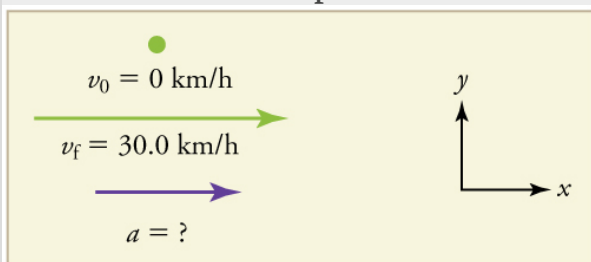
Distance is a scalar. It has magnitude but no sign to indicate direction.

Example:**Calculating Acceleration: A Subway Train Speeding Up**

Suppose the train in [\[link\]](#)(a) accelerates from rest to 30.0 km/h in the first 20.0 s of its motion. What is its average acceleration during that time interval?

Strategy

It is worth it at this point to make a simple sketch:



This problem involves three steps. First we must determine the change in velocity, then we must determine the change in time, and finally we use these values to calculate the acceleration.

Solution

1. Identify the knowns. $v_0 = 0$ (the train starts at rest), $v_f = 30.0 \text{ km/h}$, and $\Delta t = 20.0 \text{ s}$.
2. Calculate Δv . Since the train starts from rest, its change in velocity is $\Delta v = +30.0 \text{ km/h}$, where the plus sign means velocity to the right.
3. Plug in known values and solve for the unknown, \bar{a} .

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{+30.0 \text{ km/h}}{20.0 \text{ s}}$$

4. Since the units are mixed (we have both hours and seconds for time), we need to convert everything into SI units of meters and seconds. (See [Physical Quantities and Units](#) for more guidance.)

Equation:

$$\bar{a} = \left(\frac{+30 \text{ km/h}}{20.0 \text{ s}} \right) \left(\frac{10^3 \text{ m}}{1 \text{ km}} \right) \left(\frac{1 \text{ h}}{3600 \text{ s}} \right) = 0.417 \text{ m/s}^2$$

Discussion

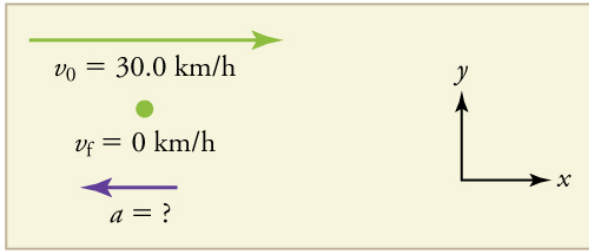
The plus sign means that acceleration is to the right. This is reasonable because the train starts from rest and ends up with a velocity to the right (also positive). So acceleration is in the same direction as the *change* in velocity, as is always the case.

Example:

Calculate Acceleration: A Subway Train Slowing Down

Now suppose that at the end of its trip, the train in [\[link\]](#)(a) slows to a stop from a speed of 30.0 km/h in 8.00 s . What is its average acceleration while stopping?

Strategy



In this case, the train is decelerating and its acceleration is negative because it is toward the left. As in the previous example, we must find the change in velocity and the change in time and then solve for acceleration.

Solution

1. Identify the knowns. $v_0 = 30.0 \text{ km/h}$, $v_f = 0 \text{ km/h}$ (the train is stopped, so its velocity is 0), and $\Delta t = 8.00 \text{ s}$.
2. Solve for the change in velocity, Δv .

Equation:

$$\Delta v = v_f - v_0 = 0 - 30.0 \text{ km/h} = -30.0 \text{ km/h}$$

3. Plug in the knowns, Δv and Δt , and solve for \bar{a} .

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{-30.0 \text{ km/h}}{8.00 \text{ s}}$$

4. Convert the units to meters and seconds.

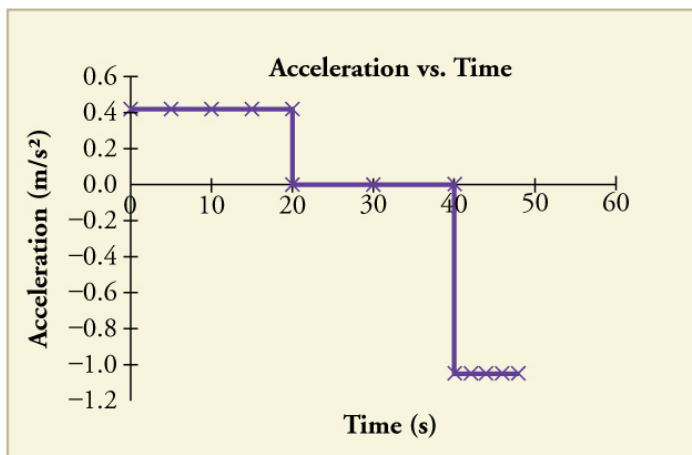
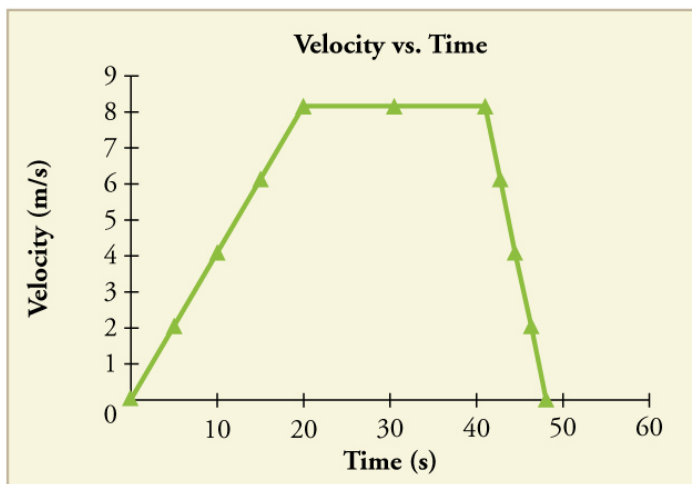
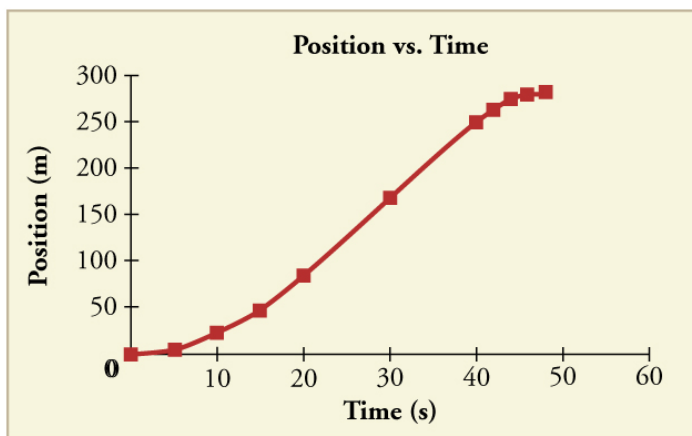
Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \left(\frac{-30.0 \text{ km/h}}{8.00 \text{ s}} \right) \left(\frac{10^3 \text{ m}}{1 \text{ km}} \right) \left(\frac{1 \text{ h}}{3600 \text{ s}} \right) = -1.04 \text{ m/s}^2.$$

Discussion

The minus sign indicates that acceleration is to the left. This sign is reasonable because the train initially has a positive velocity in this problem, and a negative acceleration would oppose the motion. Again, acceleration is in the same direction as the *change* in velocity, which is negative here. This acceleration can be called a deceleration because it has a direction opposite to the velocity.

The graphs of position, velocity, and acceleration vs. time for the trains in [\[link\]](#) and [\[link\]](#) are displayed in [\[link\]](#). (We have taken the velocity to remain constant from 20 to 40 s, after which the train decelerates.)



(a) Position of the train over time.

Notice that the train's position changes slowly at the beginning of the journey, then more and more quickly as it picks up speed. Its position then changes more slowly as it slows down at the end of the journey. In the middle of the journey, while the velocity remains constant, the position changes at a constant rate. (b) Velocity

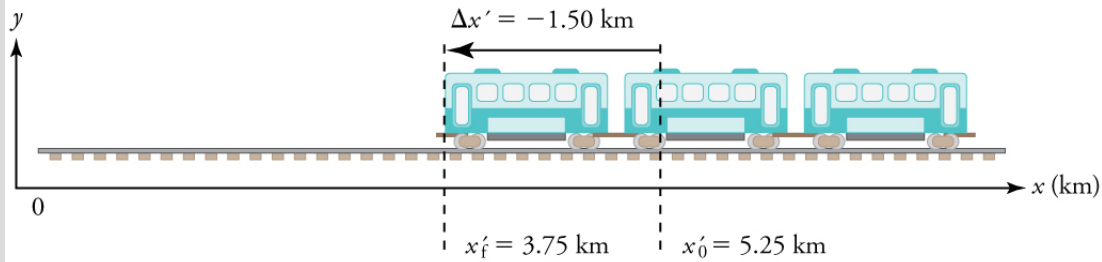
of the train over time. The train's velocity increases as it accelerates at the beginning of the journey. It remains the same in the middle of the journey (where there is no acceleration). It decreases as the train decelerates at the end of the journey.

(c) The acceleration of the train over time. The train has positive acceleration as it speeds up at the beginning of the journey. It has no acceleration as it travels at constant velocity in the middle of the journey. Its acceleration is negative as it slows down at the end of the journey.

Example:

Calculating Average Velocity: The Subway Train

What is the average velocity of the train in part b of [\[link\]](#), and shown again below, if it takes 5.00 min to make its trip?



Strategy

Average velocity is displacement divided by time. It will be negative here, since the train moves to the left and has a negative displacement.

Solution

1. Identify the knowns. $x'_f = 3.75$ km, $x'_0 = 5.25$ km, $\Delta t = 5.00$ min.
2. Determine displacement, $\Delta x'$. We found $\Delta x'$ to be -1.5 km in [\[link\]](#).
3. Solve for average velocity.

Equation:

$$\bar{v} = \frac{\Delta x'}{\Delta t} = \frac{-1.50 \text{ km}}{5.00 \text{ min}}$$

4. Convert units.

Equation:

$$\bar{v} = \frac{\Delta x'}{\Delta t} = \left(\frac{-1.50 \text{ km}}{5.00 \text{ min}} \right) \left(\frac{60 \text{ min}}{1 \text{ h}} \right) = -18.0 \text{ km/h}$$

Discussion

The negative velocity indicates motion to the left.

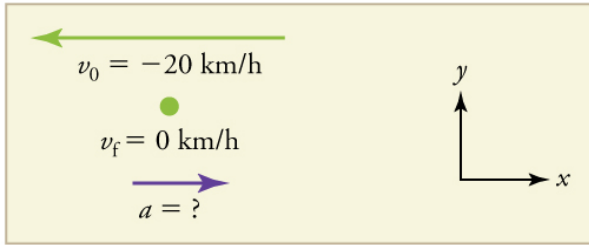
Example:

Calculating Deceleration: The Subway Train

Finally, suppose the train in [\[link\]](#) slows to a stop from a velocity of 20.0 km/h in 10.0 s. What is its average acceleration?

Strategy

Once again, let's draw a sketch:



As before, we must find the change in velocity and the change in time to calculate average acceleration.

Solution

1. Identify the knowns. $v_0 = -20 \text{ km/h}$, $v_f = 0 \text{ km/h}$, $\Delta t = 10.0 \text{ s}$.
2. Calculate Δv . The change in velocity here is actually positive, since

Equation:

$$\Delta v = v_f - v_0 = 0 - (-20 \text{ km/h}) = +20 \text{ km/h}.$$

3. Solve for \bar{a} .

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{+20.0 \text{ km/h}}{10.0 \text{ s}}$$

4. Convert units.

Equation:

$$\bar{a} = \left(\frac{+20.0 \text{ km/h}}{10.0 \text{ s}} \right) \left(\frac{10^3 \text{ m}}{1 \text{ km}} \right) \left(\frac{1 \text{ h}}{3600 \text{ s}} \right) = +0.556 \text{ m/s}^2$$

Discussion

The plus sign means that acceleration is to the right. This is reasonable because the train initially has a negative velocity (to the left) in this problem and a positive acceleration opposes the motion (and so it is to the right). Again, acceleration is in the same direction as the *change* in velocity, which is positive here. As in [\[link\]](#), this acceleration can be called a deceleration since it is in the direction opposite to the velocity.

Sign and Direction

Perhaps the most important thing to note about these examples is the signs of the answers. In our chosen coordinate system, plus means the quantity is to the right and minus means it is to the left. This is easy to imagine for displacement and velocity. But it is a little less obvious for acceleration. Most people interpret negative acceleration as the slowing of an object. This was not the case in [\[link\]](#), where a positive acceleration slowed a negative velocity. The crucial distinction was that the acceleration was in the opposite direction from the velocity. In fact, a negative acceleration will *increase* a negative velocity. For example, the train moving to the left in [\[link\]](#) is sped up by an acceleration to the left. In that case, both v and a are negative. The plus and minus signs give the directions of the accelerations. If acceleration has the same sign as the velocity, the object is speeding up. If acceleration has the opposite sign as the velocity, the object is slowing down.

Exercise:

Check Your Understanding

Problem:

An airplane lands on a runway traveling east. Describe its acceleration.

Solution:

If we take east to be positive, then the airplane has negative acceleration, as it is accelerating toward the west. It is also decelerating: its acceleration is opposite in direction to its velocity.

Note:

PhET Explorations: Moving Man Simulation

Learn about position, velocity, and acceleration graphs. Move the little man back and forth with the mouse and plot his motion. Set the position, velocity, or acceleration and let the simulation move the man for you.

<https://archive.cnx.org/specials/e2ca52af-8c6b-450e-ac2f-9300b38e8739/moving-man/>

Section Summary

- Acceleration is the rate at which velocity changes. In symbols, **average acceleration** \bar{a} is

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0}.$$

- The SI unit for acceleration is m/s^2 .
- Acceleration is a vector, and thus has both a magnitude and direction.
- Acceleration can be caused by either a change in the magnitude or the direction of the velocity.
- Instantaneous acceleration a is the acceleration at a specific instant in time.
- Deceleration is an acceleration with a direction opposite to that of the velocity.

Conceptual Questions

Exercise:

Problem:

Is it possible for speed to be constant while acceleration is not zero?
Give an example of such a situation.

Exercise:

Problem:

Is it possible for velocity to be constant while acceleration is not zero?
Explain.

Exercise:

Problem:

Give an example in which velocity is zero yet acceleration is not.

Exercise:**Problem:**

If a subway train is moving to the left (has a negative velocity) and then comes to a stop, what is the direction of its acceleration? Is the acceleration positive or negative?

Exercise:**Problem:**

Plus and minus signs are used in one-dimensional motion to indicate direction. What is the sign of an acceleration that reduces the magnitude of a negative velocity? Of a positive velocity?

Problems & Exercises**Exercise:****Problem:**

A cheetah can accelerate from rest to a speed of 30.0 m/s in 7.00 s. What is its acceleration?

Solution:

$$4.29 \text{ m/s}^2$$

Exercise:**Problem: Professional Application**

Dr. John Paul Stapp was U.S. Air Force officer who studied the effects of extreme deceleration on the human body. On December 10, 1954, Stapp rode a rocket sled, accelerating from rest to a top speed of 282 m/s (1015 km/h) in 5.00 s, and was brought jarringly back to rest in only 1.40 s! Calculate his (a) acceleration and (b) deceleration.

Express each in multiples of g (9.80 m/s^2) by taking its ratio to the acceleration of gravity.

Exercise:

Problem:

A commuter backs her car out of her garage with an acceleration of 1.40 m/s^2 . (a) How long does it take her to reach a speed of 2.00 m/s ? (b) If she then brakes to a stop in 0.800 s , what is her deceleration?

Solution:

(a) 1.43 s

(b) -2.50 m/s^2

Exercise:

Problem:

Assume that an intercontinental ballistic missile goes from rest to a suborbital speed of 6.50 km/s in 60.0 s (the actual speed and time are classified). What is its average acceleration in m/s^2 and in multiples of g (9.80 m/s^2)?

Glossary

acceleration

the rate of change in velocity; the change in velocity over time

average acceleration

the change in velocity divided by the time over which it changes

instantaneous acceleration

acceleration at a specific point in time

deceleration

acceleration in the direction opposite to velocity; acceleration that results in a decrease in velocity

Motion Equations for Constant Acceleration in One Dimension

- Calculate displacement of an object that is not accelerating, given initial position and velocity.
- Calculate final velocity of an accelerating object, given initial velocity, acceleration, and time.
- Calculate displacement and final position of an accelerating object, given initial position, initial velocity, time, and acceleration.



Kinematic equations can help us describe and predict the motion of moving objects such as these kayaks racing in Newbury, England. (credit: Barry Skeates, Flickr)

We might know that the greater the acceleration of, say, a car moving away from a stop sign, the greater the displacement in a given time. But we have not developed a specific equation that relates acceleration and displacement. In this section, we develop some convenient equations for kinematic relationships, starting from the definitions of displacement, velocity, and acceleration already covered.

Notation: t , x , v , a

First, let us make some simplifications in notation. Taking the initial time to be zero, as if time is measured with a stopwatch, is a great simplification. Since elapsed time is $\Delta t = t_f - t_0$, taking $t_0 = 0$ means that $\Delta t = t_f$, the final time on the stopwatch. When initial time is taken to be zero, we use the subscript 0 to denote initial values of position and velocity. That is, x_0 is *the initial position* and v_0 is *the initial velocity*. We put no subscripts on the final values. That is, t is *the final time*, x is *the final position*, and v is *the final velocity*. This gives a simpler expression for elapsed time—now, $\Delta t = t$. It also simplifies the expression for displacement, which is now $\Delta x = x - x_0$. Also, it simplifies the expression for change in velocity, which is now $\Delta v = v - v_0$. To summarize, using the simplified notation, with the initial time taken to be zero,

Equation:

$$\begin{aligned}\Delta t &= t \\ \Delta x &= x - x_0 \\ \Delta v &= v - v_0\end{aligned}$$

where *the subscript 0 denotes an initial value and the absence of a subscript denotes a final value* in whatever motion is under consideration.

We now make the important assumption that *acceleration is constant*. This assumption allows us to avoid using calculus to find instantaneous acceleration. Since acceleration is constant, the average and instantaneous accelerations are equal. That is,

Equation:

$$\bar{a} = a = \text{constant},$$

so we use the symbol a for acceleration at all times. Assuming acceleration to be constant does not seriously limit the situations we can study nor degrade the accuracy of our treatment. For one thing, acceleration *is* constant in a great number of situations. Furthermore, in many other situations we can accurately describe motion by assuming a constant acceleration equal to the average acceleration for that motion. Finally, in

motions where acceleration changes drastically, such as a car accelerating to top speed and then braking to a stop, the motion can be considered in separate parts, each of which has its own constant acceleration.

Note:

Solving for Displacement (Δx) and Final Position (x) from Average Velocity when Acceleration (a) is Constant

To get our first two new equations, we start with the definition of average velocity:

Equation:

$$\bar{v} = \frac{\Delta x}{\Delta t}.$$

Substituting the simplified notation for Δx and Δt yields

Equation:

$$\bar{v} = \frac{x - x_0}{t}.$$

Solving for x yields

Equation:

$$x = x_0 + \bar{v}t,$$

where the average velocity is

Equation:

$$\bar{v} = \frac{v_0 + v}{2} \text{ (constant } a\text{)}.$$

The equation $\bar{v} = \frac{v_0 + v}{2}$ reflects the fact that, when acceleration is constant, \bar{v} is just the simple average of the initial and final velocities. For example, if

you steadily increase your velocity (that is, with constant acceleration) from 30 to 60 km/h, then your average velocity during this steady increase is 45 km/h. Using the equation $\bar{v} = \frac{v_0 + v}{2}$ to check this, we see that

Equation:

$$\bar{v} = \frac{v_0 + v}{2} = \frac{30 \text{ km/h} + 60 \text{ km/h}}{2} = 45 \text{ km/h},$$

which seems logical.

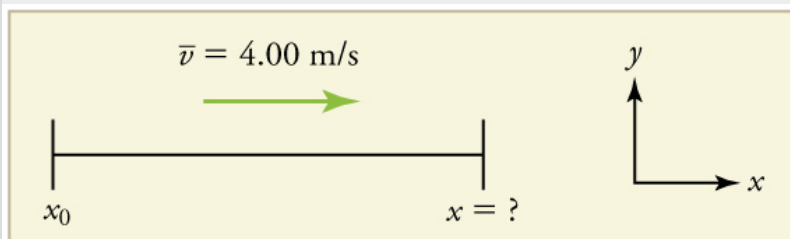
Example:

Calculating Displacement: How Far does the Jogger Run?

A jogger runs down a straight stretch of road with an average velocity of 4.00 m/s for 2.00 min. What is his final position, taking his initial position to be zero?

Strategy

Draw a sketch.



The final position x is given by the equation

Equation:

$$x = x_0 + \bar{v}t.$$

To find x , we identify the values of x_0 , \bar{v} , and t from the statement of the problem and substitute them into the equation.

Solution

1. Identify the knowns. $\bar{v} = 4.00 \text{ m/s}$, $\Delta t = 2.00 \text{ min}$, and $x_0 = 0 \text{ m}$.
2. Enter the known values into the equation.

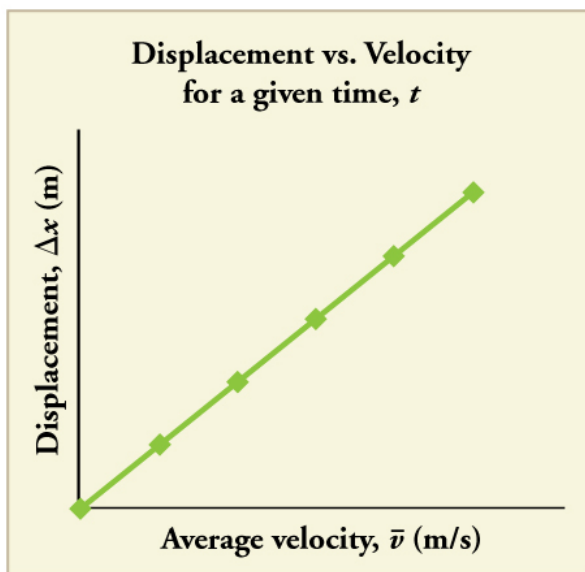
Equation:

$$x = x_0 + \bar{v}t = 0 + (4.00 \text{ m/s})(120 \text{ s}) = 480 \text{ m}$$

Discussion

Velocity and final displacement are both positive, which means they are in the same direction.

The equation $x = x_0 + \bar{v}t$ gives insight into the relationship between displacement, average velocity, and time. It shows, for example, that displacement is a linear function of average velocity. (By linear function, we mean that displacement depends on \bar{v} rather than on \bar{v} raised to some other power, such as \bar{v}^2 . When graphed, linear functions look like straight lines with a constant slope.) On a car trip, for example, we will get twice as far in a given time if we average 90 km/h than if we average 45 km/h.



There is a linear relationship between displacement and average velocity. For a given time t , an object moving twice as fast as another object will

move twice as far as the other object.

Note:

Solving for Final Velocity

We can derive another useful equation by manipulating the definition of acceleration.

Equation:

$$a = \frac{\Delta v}{\Delta t}$$

Substituting the simplified notation for Δv and Δt gives us

Equation:

$$a = \frac{v - v_0}{t} \text{ (constant } a\text{)}.$$

Solving for v yields

Equation:

$$v = v_0 + at \text{ (constant } a\text{)}.$$

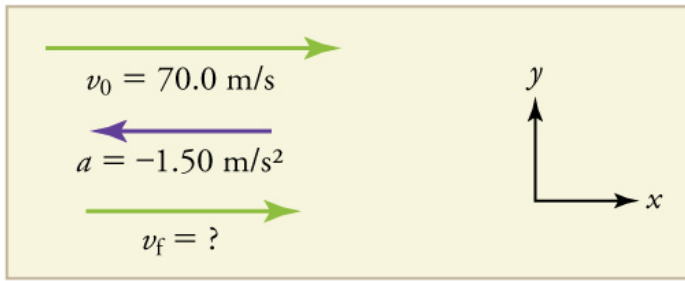
Example:

Calculating Final Velocity: An Airplane Slowing Down after Landing

An airplane lands with an initial velocity of 70.0 m/s and then decelerates at 1.50 m/s² for 40.0 s. What is its final velocity?

Strategy

Draw a sketch. We draw the acceleration vector in the direction opposite the velocity vector because the plane is decelerating.



Solution

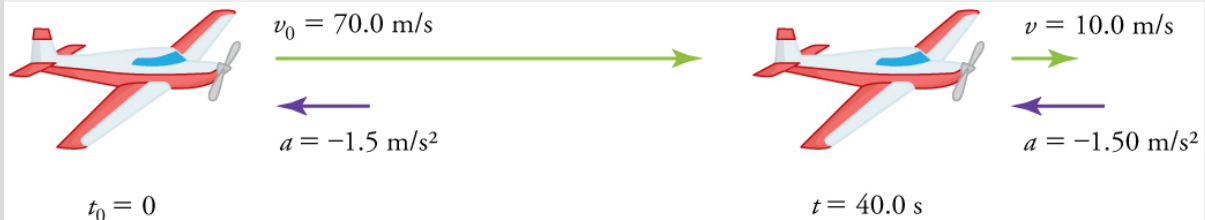
1. Identify the knowns. $v_0 = 70.0 \text{ m/s}$, $a = -1.50 \text{ m/s}^2$, $t = 40.0 \text{ s}$.
2. Identify the unknown. In this case, it is final velocity, v_f .
3. Determine which equation to use. We can calculate the final velocity using the equation $v = v_0 + at$.
4. Plug in the known values and solve.

Equation:

$$v = v_0 + at = 70.0 \text{ m/s} + (-1.50 \text{ m/s}^2)(40.0 \text{ s}) = 10.0 \text{ m/s}$$

Discussion

The final velocity is much less than the initial velocity, as desired when slowing down, but still positive. With jet engines, reverse thrust could be maintained long enough to stop the plane and start moving it backward. That would be indicated by a negative final velocity, which is not the case here.



The airplane lands with an initial velocity of 70.0 m/s and slows to a final velocity of 10.0 m/s before heading for the terminal. Note that the acceleration is negative because its direction is opposite to its velocity, which is positive.

In addition to being useful in problem solving, the equation $v = v_0 + at$ gives us insight into the relationships among velocity, acceleration, and time. From it we can see, for example, that

- final velocity depends on how large the acceleration is and how long it lasts
- if the acceleration is zero, then the final velocity equals the initial velocity ($v = v_0$), as expected (i.e., velocity is constant)
- if a is negative, then the final velocity is less than the initial velocity

(All of these observations fit our intuition, and it is always useful to examine basic equations in light of our intuition and experiences to check that they do indeed describe nature accurately.)

Note:

Making Connections: Real-World Connection



The Space Shuttle *Endeavor*
blasts off from the Kennedy
Space Center in February 2010.
(credit: Matthew Simantov,
Flickr)

An intercontinental ballistic missile (ICBM) has a larger average acceleration than the Space Shuttle and achieves a greater velocity in the

first minute or two of flight (actual ICBM burn times are classified—short-burn-time missiles are more difficult for an enemy to destroy). But the Space Shuttle obtains a greater final velocity, so that it can orbit the earth rather than come directly back down as an ICBM does. The Space Shuttle does this by accelerating for a longer time.

Note:

Solving for Final Position When Velocity is Not Constant ($a \neq 0$)

We can combine the equations above to find a third equation that allows us to calculate the final position of an object experiencing constant acceleration. We start with

Equation:

$$v = v_0 + at.$$

Adding v_0 to each side of this equation and dividing by 2 gives

Equation:

$$\frac{v_0 + v}{2} = v_0 + \frac{1}{2}at.$$

Since $\frac{v_0 + v}{2} = \bar{v}$ for constant acceleration, then

Equation:

$$\bar{v} = v_0 + \frac{1}{2}at.$$

Now we substitute this expression for \bar{v} into the equation for displacement,

$x = x_0 + \bar{v}t$, yielding

Equation:

$$x = x_0 + v_0t + \frac{1}{2}at^2 \text{ (constant } a\text{)}.$$

Example:

Calculating Displacement of an Accelerating Object: Dragsters

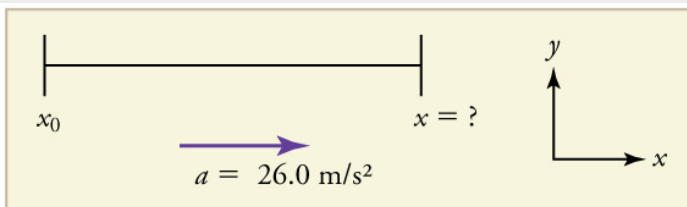
Dragsters can achieve average accelerations of 26.0 m/s^2 . Suppose such a dragster accelerates from rest at this rate for 5.56 s . How far does it travel in this time?



U.S. Army Top Fuel pilot
Tony “The Sarge”
Schumacher begins a race
with a controlled burnout.
(credit: Lt. Col. William
Thurmond. Photo
Courtesy of U.S. Army.)

Strategy

Draw a sketch.



We are asked to find displacement, which is x if we take x_0 to be zero. (Think about it like the starting line of a race. It can be anywhere, but we call it 0 and measure all other positions relative to it.) We can use the equation $x = x_0 + v_0t + \frac{1}{2}at^2$ once we identify v_0 , a , and t from the statement of the problem.

Solution

1. Identify the knowns. Starting from rest means that $v_0 = 0$, a is given as 26.0 m/s^2 and t is given as 5.56 s .

2. Plug the known values into the equation to solve for the unknown x :

Equation:

$$x = x_0 + v_0 t + \frac{1}{2} a t^2.$$

Since the initial position and velocity are both zero, this simplifies to

Equation:

$$x = \frac{1}{2} a t^2.$$

Substituting the identified values of a and t gives

Equation:

$$x = \frac{1}{2} (26.0 \text{ m/s}^2) (5.56 \text{ s})^2,$$

yielding

Equation:

$$x = 402 \text{ m}.$$

Discussion

If we convert 402 m to miles, we find that the distance covered is very close to one quarter of a mile, the standard distance for drag racing. So the answer is reasonable. This is an impressive displacement in only 5.56 s , but top-notch dragsters can do a quarter mile in even less time than this.

What else can we learn by examining the equation $x = x_0 + v_0 t + \frac{1}{2} a t^2$? We see that:

- displacement depends on the square of the elapsed time when acceleration is not zero. In [\[link\]](#), the dragster covers only one fourth of the total distance in the first half of the elapsed time

- if acceleration is zero, then the initial velocity equals average velocity ($v_0 = \bar{v}$) and $x = x_0 + v_0 t + \frac{1}{2}at^2$ becomes $x = x_0 + v_0 t$

Note:

Solving for Final Velocity when Velocity Is Not Constant ($a \neq 0$)

A fourth useful equation can be obtained from another algebraic manipulation of previous equations.

If we solve $v = v_0 + at$ for t , we get

Equation:

$$t = \frac{v - v_0}{a}.$$

Substituting this and $\bar{v} = \frac{v_0 + v}{2}$ into $x = x_0 + \bar{v}t$, we get

Equation:

$$v^2 = v_0^2 + 2a(x - x_0) \text{ (constant } a\text{)}.$$

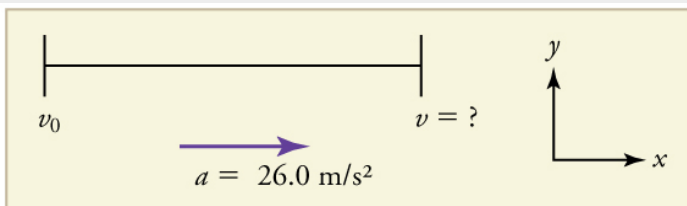
Example:

Calculating Final Velocity: Dragsters

Calculate the final velocity of the dragster in [\[link\]](#) without using information about time.

Strategy

Draw a sketch.



The equation $v^2 = v_0^2 + 2a(x - x_0)$ is ideally suited to this task because it relates velocities, acceleration, and displacement, and no time information is required.

Solution

1. Identify the known values. We know that $v_0 = 0$, since the dragster starts from rest. Then we note that $x - x_0 = 402$ m (this was the answer in [\[link\]](#)). Finally, the average acceleration was given to be $a = 26.0$ m/s².
2. Plug the knowns into the equation $v^2 = v_0^2 + 2a(x - x_0)$ and solve for v .

Equation:

$$v^2 = 0 + 2(26.0 \text{ m/s}^2)(402 \text{ m}).$$

Thus

Equation:

$$v^2 = 2.09 \times 10^4 \text{ m}^2/\text{s}^2.$$

To get v , we take the square root:

Equation:

$$v = \sqrt{2.09 \times 10^4 \text{ m}^2/\text{s}^2} = 145 \text{ m/s}.$$

Discussion

145 m/s is about 522 km/h or about 324 mi/h, but even this breakneck speed is short of the record for the quarter mile. Also, note that a square root has two values; we took the positive value to indicate a velocity in the same direction as the acceleration.

An examination of the equation $v^2 = v_0^2 + 2a(x - x_0)$ can produce further insights into the general relationships among physical quantities:

- The final velocity depends on how large the acceleration is and the distance over which it acts
- For a fixed deceleration, a car that is going twice as fast doesn't simply stop in twice the distance—it takes much further to stop. (This is why

we have reduced speed zones near schools.)

Putting Equations Together

In the following examples, we further explore one-dimensional motion, but in situations requiring slightly more algebraic manipulation. The examples also give insight into problem-solving techniques. The box below provides easy reference to the equations needed.

Note:

Summary of Kinematic Equations (constant a)

Equation:

$$x = x_0 + \bar{v}t$$

Equation:

$$\bar{v} = \frac{v_0 + v}{2}$$

Equation:

$$v = v_0 + at$$

Equation:

$$x = x_0 + v_0t + \frac{1}{2}at^2$$

Equation:

$$v^2 = v_0^2 + 2a(x - x_0)$$

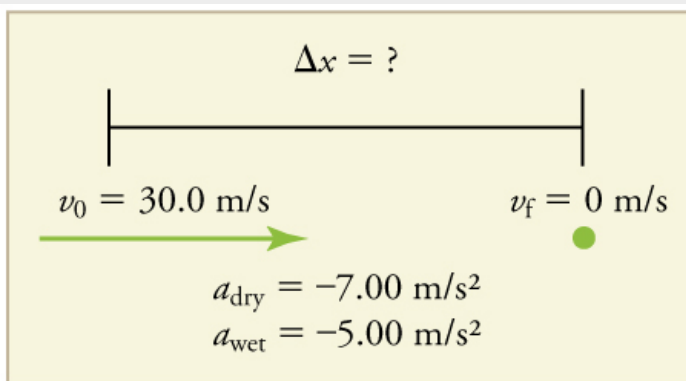
Example:

Calculating Displacement: How Far Does a Car Go When Coming to a Halt?

On dry concrete, a car can decelerate at a rate of 7.00 m/s^2 , whereas on wet concrete it can decelerate at only 5.00 m/s^2 . Find the distances necessary to stop a car moving at 30.0 m/s (about 110 km/h) (a) on dry concrete and (b) on wet concrete. (c) Repeat both calculations, finding the displacement from the point where the driver sees a traffic light turn red, taking into account his reaction time of 0.500 s to get his foot on the brake.

Strategy

Draw a sketch.



In order to determine which equations are best to use, we need to list all of the known values and identify exactly what we need to solve for. We shall do this explicitly in the next several examples, using tables to set them off.

Solution for (a)

1. Identify the knowns and what we want to solve for. We know that $v_0 = 30.0 \text{ m/s}$; $v = 0$; $a = -7.00 \text{ m/s}^2$ (a is negative because it is in a direction opposite to velocity). We take x_0 to be 0. We are looking for displacement Δx , or $x - x_0$.

2. Identify the equation that will help up solve the problem. The best equation to use is

Equation:

$$v^2 = v_0^2 + 2a(x - x_0).$$

This equation is best because it includes only one unknown, x . We know the values of all the other variables in this equation. (There are other equations that would allow us to solve for x , but they require us to know

the stopping time, t , which we do not know. We could use them but it would entail additional calculations.)

3. Rearrange the equation to solve for x .

Equation:

$$x - x_0 = \frac{v^2 - v_0^2}{2a}$$

4. Enter known values.

Equation:

$$x - 0 = \frac{0^2 - (30.0 \text{ m/s})^2}{2(-7.00 \text{ m/s}^2)}$$

Thus,

Equation:

$$x = 64.3 \text{ m on dry concrete.}$$

Solution for (b)

This part can be solved in exactly the same manner as Part A. The only difference is that the deceleration is -5.00 m/s^2 . The result is

Equation:

$$x_{\text{wet}} = 90.0 \text{ m on wet concrete.}$$

Solution for (c)

Once the driver reacts, the stopping distance is the same as it is in Parts A and B for dry and wet concrete. So to answer this question, we need to calculate how far the car travels during the reaction time, and then add that to the stopping time. It is reasonable to assume that the velocity remains constant during the driver's reaction time.

1. Identify the knowns and what we want to solve for. We know that

$\bar{v} = 30.0 \text{ m/s}$; $t_{\text{reaction}} = 0.500 \text{ s}$; $a_{\text{reaction}} = 0$. We take $x_{0-\text{reaction}}$ to be 0. We are looking for x_{reaction} .

2. Identify the best equation to use.

$x = x_0 + \bar{v}t$ works well because the only unknown value is x , which is what we want to solve for.

3. Plug in the knowns to solve the equation.

Equation:

$$x = 0 + (30.0 \text{ m/s})(0.500 \text{ s}) = 15.0 \text{ m}.$$

This means the car travels 15.0 m while the driver reacts, making the total displacements in the two cases of dry and wet concrete 15.0 m greater than if he reacted instantly.

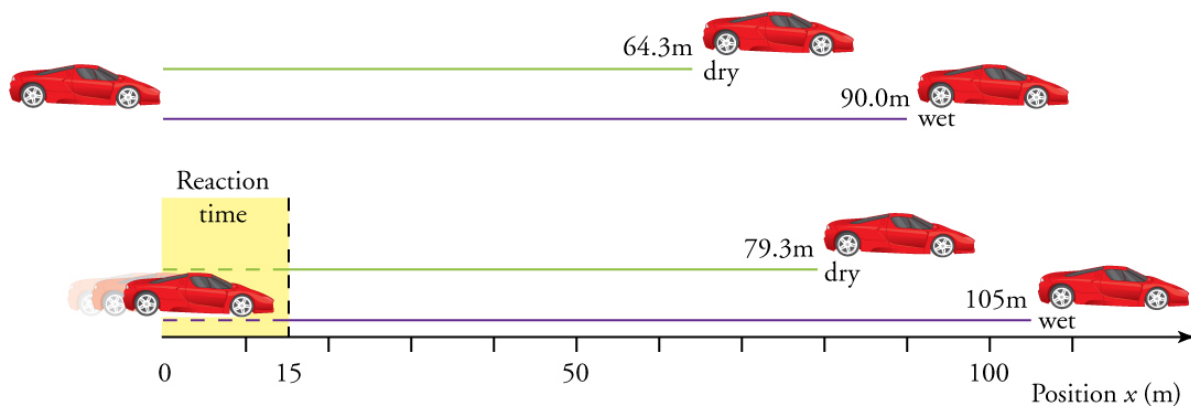
4. Add the displacement during the reaction time to the displacement when braking.

Equation:

$$x_{\text{braking}} + x_{\text{reaction}} = x_{\text{total}}$$

a. $64.3 \text{ m} + 15.0 \text{ m} = 79.3 \text{ m}$ when dry

b. $90.0 \text{ m} + 15.0 \text{ m} = 105 \text{ m}$ when wet



The distance necessary to stop a car varies greatly, depending on road conditions and driver reaction time. Shown here are the braking distances for dry and wet pavement, as calculated in this example, for a car initially traveling at 30.0 m/s. Also shown are the total distances traveled from the point where the driver first sees a light turn red, assuming a 0.500 s reaction time.

Discussion

The displacements found in this example seem reasonable for stopping a fast-moving car. It should take longer to stop a car on wet rather than dry pavement. It is interesting that reaction time adds significantly to the displacements. But more important is the general approach to solving problems. We identify the knowns and the quantities to be determined and then find an appropriate equation. There is often more than one way to solve a problem. The various parts of this example can in fact be solved by other methods, but the solutions presented above are the shortest.

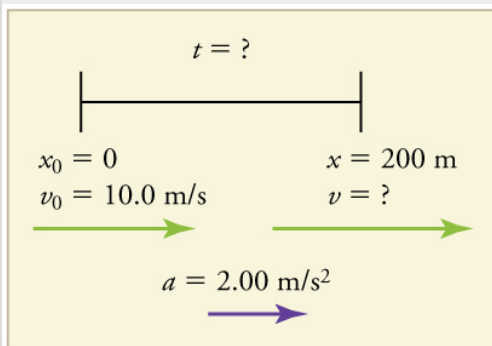
Example:

Calculating Time: A Car Merges into Traffic

Suppose a car merges into freeway traffic on a 200-m-long ramp. If its initial velocity is 10.0 m/s and it accelerates at 2.00 m/s^2 , how long does it take to travel the 200 m up the ramp? (Such information might be useful to a traffic engineer.)

Strategy

Draw a sketch.



We are asked to solve for the time t . As before, we identify the known quantities in order to choose a convenient physical relationship (that is, an equation with one unknown, t).

Solution

1. Identify the knowns and what we want to solve for. We know that $v_0 = 10 \text{ m/s}$; $a = 2.00 \text{ m/s}^2$; and $x = 200 \text{ m}$.
2. We need to solve for t . Choose the best equation. $x = x_0 + v_0 t + \frac{1}{2} a t^2$ works best because the only unknown in the equation is the variable t for which we need to solve.

3. We will need to rearrange the equation to solve for t . In this case, it will be easier to plug in the knowns first.

Equation:

$$200 \text{ m} = 0 \text{ m} + (10.0 \text{ m/s})t + \frac{1}{2} (2.00 \text{ m/s}^2) t^2$$

4. Simplify the equation. The units of meters (m) cancel because they are in each term. We can get the units of seconds (s) to cancel by taking $t = t \text{ s}$, where t is the magnitude of time and s is the unit. Doing so leaves

Equation:

$$200 = 10t + t^2.$$

5. Use the quadratic formula to solve for t .

(a) Rearrange the equation to get 0 on one side of the equation.

Equation:

$$t^2 + 10t - 200 = 0$$

This is a quadratic equation of the form

Equation:

$$at^2 + bt + c = 0,$$

where the constants are $a = 1.00$, $b = 10.0$, and $c = -200$.

(b) Its solutions are given by the quadratic formula:

Equation:

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

This yields two solutions for t , which are

Equation:

$$t = 10.0 \text{ and } -20.0.$$

In this case, then, the time is $t = t$ in seconds, or

Equation:

$$t = 10.0 \text{ s and } -20.0 \text{ s.}$$

A negative value for time is unreasonable, since it would mean that the event happened 20 s before the motion began. We can discard that solution. Thus,

Equation:

$$t = 10.0 \text{ s.}$$

Discussion

Whenever an equation contains an unknown squared, there will be two solutions. In some problems both solutions are meaningful, but in others, such as the above, only one solution is reasonable. The 10.0 s answer seems reasonable for a typical freeway on-ramp.

With the basics of kinematics established, we can go on to many other interesting examples and applications. In the process of developing kinematics, we have also glimpsed a general approach to problem solving that produces both correct answers and insights into physical relationships. [Problem-Solving Basics](#) discusses problem-solving basics and outlines an approach that will help you succeed in this invaluable task.

Note:

Making Connections: Take-Home Experiment—Breaking News

We have been using SI units of meters per second squared to describe some examples of acceleration or deceleration of cars, runners, and trains. To achieve a better feel for these numbers, one can measure the braking deceleration of a car doing a slow (and safe) stop. Recall that, for average acceleration, $\bar{a} = \Delta v / \Delta t$. While traveling in a car, slowly apply the brakes as you come up to a stop sign. Have a passenger note the initial speed in miles per hour and the time taken (in seconds) to stop. From this, calculate the deceleration in miles per hour per second. Convert this to meters per second squared and compare with other decelerations mentioned in this chapter. Calculate the distance traveled in braking.

Exercise:
Check Your Understanding

Problem:

A manned rocket accelerates at a rate of 20 m/s^2 during launch. How long does it take the rocket to reach a velocity of 400 m/s ?

Solution:

To answer this, choose an equation that allows you to solve for time t , given only a , v_0 , and v .

Equation:

$$v = v_0 + at$$

Rearrange to solve for t .

Equation:

$$t = \frac{v - v_0}{a} = \frac{400 \text{ m/s} - 0 \text{ m/s}}{20 \text{ m/s}^2} = 20 \text{ s}$$

Section Summary

- To simplify calculations we take acceleration to be constant, so that $\bar{a} = a$ at all times.
- We also take initial time to be zero.
- Initial position and velocity are given a subscript 0; final values have no subscript. Thus,

Equation:

$$\Delta t = t$$

$$\Delta x = x - x_0$$

$$\Delta v = v - v_0$$

- The following kinematic equations for motion with constant a are useful:

Equation:

$$x = x_0 + \bar{v}t$$

Equation:

$$\bar{v} = \frac{v_0 + v}{2}$$

Equation:

$$v = v_0 + at$$

Equation:

$$x = x_0 + v_0t + \frac{1}{2}at^2$$

Equation:

$$v^2 = v_0^2 + 2a(x - x_0)$$

- In vertical motion, y is substituted for x .

Problems & Exercises

Exercise:

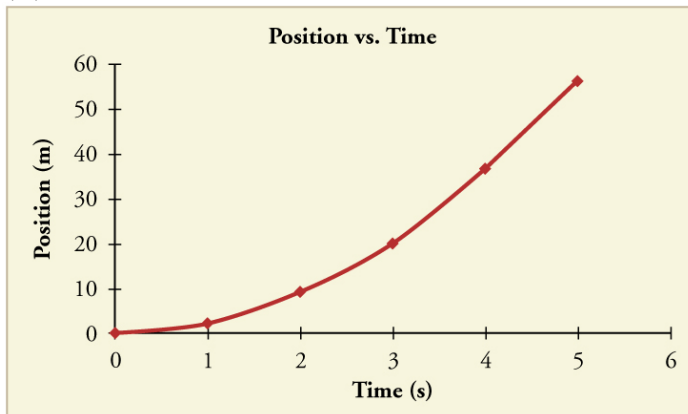
Problem:

An Olympic-class sprinter starts a race with an acceleration of 4.50 m/s^2 . (a) What is her speed 2.40 s later? (b) Sketch a graph of her position vs. time for this period.

Solution:

(a) 10.8 m/s

(b)



Exercise:

Problem:

A well-thrown ball is caught in a well-padded mitt. If the deceleration of the ball is $2.10 \times 10^4 \text{ m/s}^2$, and 1.85 ms ($1 \text{ ms} = 10^{-3} \text{ s}$) elapses from the time the ball first touches the mitt until it stops, what was the initial velocity of the ball?

Solution:

38.9 m/s (about 87 miles per hour)

Exercise:

Problem:

A bullet in a gun is accelerated from the firing chamber to the end of the barrel at an average rate of $6.20 \times 10^5 \text{ m/s}^2$ for $8.10 \times 10^{-4} \text{ s}$. What is its muzzle velocity (that is, its final velocity)?

Exercise:

Problem:

(a) A light-rail commuter train accelerates at a rate of 1.35 m/s^2 . How long does it take to reach its top speed of 80.0 km/h , starting from rest? (b) The same train ordinarily decelerates at a rate of 1.65 m/s^2 . How long does it take to come to a stop from its top speed? (c) In emergencies the train can decelerate more rapidly, coming to rest from 80.0 km/h in 8.30 s . What is its emergency deceleration in m/s^2 ?

Solution:

(a) 16.5 s

(b) 13.5 s

(c) -2.68 m/s^2

Exercise:**Problem:**

While entering a freeway, a car accelerates from rest at a rate of 2.40 m/s^2 for 12.0 s . (a) Draw a sketch of the situation. (b) List the knowns in this problem. (c) How far does the car travel in those 12.0 s ? To solve this part, first identify the unknown, and then discuss how you chose the appropriate equation to solve for it. After choosing the equation, show your steps in solving for the unknown, check your units, and discuss whether the answer is reasonable. (d) What is the car's final velocity? Solve for this unknown in the same manner as in part (c), showing all steps explicitly.

Exercise:**Problem:**

At the end of a race, a runner decelerates from a velocity of 9.00 m/s at a rate of 2.00 m/s^2 . (a) How far does she travel in the next 5.00 s ? (b) What is her final velocity? (c) Evaluate the result. Does it make sense?

Solution:

(a) 20.0 m

(b) -1.00 m/s

(c) This result does not really make sense. If the runner starts at 9.00 m/s and decelerates at 2.00 m/s^2 , then she will have stopped after 4.50 s . If she continues to decelerate, she will be running backwards.

Exercise:**Problem: Professional Application:**

Blood is accelerated from rest to 30.0 cm/s in a distance of 1.80 cm by the left ventricle of the heart. (a) Make a sketch of the situation. (b) List the knowns in this problem. (c) How long does the acceleration take? To solve this part, first identify the unknown, and then discuss how you chose the appropriate equation to solve for it. After choosing the equation, show your steps in solving for the unknown, checking your units. (d) Is the answer reasonable when compared with the time for a heartbeat?

Exercise:**Problem:**

In a slap shot, a hockey player accelerates the puck from a velocity of 8.00 m/s to 40.0 m/s in the same direction. If this shot takes $3.33 \times 10^{-2} \text{ s}$, calculate the distance over which the puck accelerates.

Solution:

0.799 m

Exercise:

Problem:

A powerful motorcycle can accelerate from rest to 26.8 m/s (100 km/h) in only 3.90 s. (a) What is its average acceleration? (b) How far does it travel in that time?

Exercise:**Problem:**

Freight trains can produce only relatively small accelerations and decelerations. (a) What is the final velocity of a freight train that accelerates at a rate of 0.0500 m/s^2 for 8.00 min, starting with an initial velocity of 4.00 m/s? (b) If the train can slow down at a rate of 0.550 m/s^2 , how long will it take to come to a stop from this velocity? (c) How far will it travel in each case?

Solution:

(a) 28.0 m/s

(b) 50.9 s

(c) 7.68 km to accelerate and 713 m to decelerate

Exercise:**Problem:**

A fireworks shell is accelerated from rest to a velocity of 65.0 m/s over a distance of 0.250 m. (a) How long did the acceleration last? (b) Calculate the acceleration.

Exercise:

Problem:

A swan on a lake gets airborne by flapping its wings and running on top of the water. (a) If the swan must reach a velocity of 6.00 m/s to take off and it accelerates from rest at an average rate of 0.350 m/s^2 , how far will it travel before becoming airborne? (b) How long does this take?

Solution:

(a) 51.4 m

(b) 17.1 s

Exercise:**Problem: Professional Application:**

A woodpecker's brain is specially protected from large decelerations by tendon-like attachments inside the skull. While pecking on a tree, the woodpecker's head comes to a stop from an initial velocity of 0.600 m/s in a distance of only 2.00 mm. (a) Find the acceleration in m/s^2 and in multiples of g ($g = 9.80 \text{ m/s}^2$). (b) Calculate the stopping time. (c) The tendons cradling the brain stretch, making its stopping distance 4.50 mm (greater than the head and, hence, less deceleration of the brain). What is the brain's deceleration, expressed in multiples of g ?

Exercise:**Problem:**

An unwary football player collides with a padded goalpost while running at a velocity of 7.50 m/s and comes to a full stop after compressing the padding and his body 0.350 m. (a) What is his deceleration? (b) How long does the collision last?

Solution:

(a) -80.4 m/s^2

(b) $9.33 \times 10^{-2} \text{ s}$

Exercise:

Problem:

In World War II, there were several reported cases of airmen who jumped from their flaming airplanes with no parachute to escape certain death. Some fell about 20,000 feet (6000 m), and some of them survived, with few life-threatening injuries. For these lucky pilots, the tree branches and snow drifts on the ground allowed their deceleration to be relatively small. If we assume that a pilot's speed upon impact was 123 mph (54 m/s), then what was his deceleration? Assume that the trees and snow stopped him over a distance of 3.0 m.

Exercise:

Problem:

Consider a grey squirrel falling out of a tree to the ground. (a) If we ignore air resistance in this case (only for the sake of this problem), determine a squirrel's velocity just before hitting the ground, assuming it fell from a height of 3.0 m. (b) If the squirrel stops in a distance of 2.0 cm through bending its limbs, compare its deceleration with that of the airman in the previous problem.

Solution:

(a) 7.7 m/s

(b) $-15 \times 10^2 \text{ m/s}^2$. This is about 3 times the deceleration of the pilots, who were falling from thousands of meters high!

Exercise:

Problem:

An express train passes through a station. It enters with an initial velocity of 22.0 m/s and decelerates at a rate of 0.150 m/s^2 as it goes through. The station is 210 m long. (a) How long is the nose of the train in the station? (b) How fast is it going when the nose leaves the station? (c) If the train is 130 m long, when does the end of the train leave the station? (d) What is the velocity of the end of the train as it leaves?

Exercise:**Problem:**

Dragsters can actually reach a top speed of 145 m/s in only 4.45 s —considerably less time than given in [\[link\]](#) and [\[link\]](#). (a) Calculate the average acceleration for such a dragster. (b) Find the final velocity of this dragster starting from rest and accelerating at the rate found in (a) for 402 m (a quarter mile) without using any information on time. (c) Why is the final velocity greater than that used to find the average acceleration? *Hint:* Consider whether the assumption of constant acceleration is valid for a dragster. If not, discuss whether the acceleration would be greater at the beginning or end of the run and what effect that would have on the final velocity.

Solution:

(a) 32.6 m/s^2

(b) 162 m/s

(c) $v > v_{\text{max}}$, because the assumption of constant acceleration is not valid for a dragster. A dragster changes gears, and would have a greater acceleration in first gear than second gear than third gear, etc. The acceleration would be greatest at the beginning, so it would not be accelerating at 32.6 m/s^2 during the last few meters, but substantially less, and the final velocity would be less than 162 m/s .

Exercise:**Problem:**

A bicycle racer sprints at the end of a race to clinch a victory. The racer has an initial velocity of 11.5 m/s and accelerates at the rate of 0.500 m/s^2 for 7.00 s . (a) What is his final velocity? (b) The racer continues at this velocity to the finish line. If he was 300 m from the finish line when he started to accelerate, how much time did he save? (c) One other racer was 5.00 m ahead when the winner started to accelerate, but he was unable to accelerate, and traveled at 11.8 m/s until the finish line. How far ahead of him (in meters and in seconds) did the winner finish?

Exercise:**Problem:**

In 1967, New Zealander Burt Munro set the world record for an Indian motorcycle, on the Bonneville Salt Flats in Utah, with a maximum speed of 183.58 mi/h . The one-way course was 5.00 mi long. Acceleration rates are often described by the time it takes to reach 60.0 mi/h from rest. If this time was 4.00 s , and Burt accelerated at this rate until he reached his maximum speed, how long did it take Burt to complete the course?

Solution:

104 s

Exercise:

Problem:

(a) A world record was set for the men's 100-m dash in the 2008 Olympic Games in Beijing by Usain Bolt of Jamaica. Bolt "coasted" across the finish line with a time of 9.69 s. If we assume that Bolt accelerated for 3.00 s to reach his maximum speed, and maintained that speed for the rest of the race, calculate his maximum speed and his acceleration. (b) During the same Olympics, Bolt also set the world record in the 200-m dash with a time of 19.30 s. Using the same assumptions as for the 100-m dash, what was his maximum speed for this race?

Solution:

(a) $v = 12.2 \text{ m/s}$; $a = 4.07 \text{ m/s}^2$

(b) $v = 11.2 \text{ m/s}$

Problem-Solving Basics for One-Dimensional Kinematics

- Apply problem-solving steps and strategies to solve problems of one-dimensional kinematics.
- Apply strategies to determine whether or not the result of a problem is reasonable, and if not, determine the cause.



Problem-solving skills are essential to your success in Physics. (credit: scui3asteveo, Flickr)

Problem-solving skills are obviously essential to success in a quantitative course in physics. More importantly, the ability to apply broad physical principles, usually represented by equations, to specific situations is a very powerful form of knowledge. It is much more powerful than memorizing a list of facts. Analytical skills and problem-solving abilities can be applied to new situations, whereas a list of facts cannot be made long enough to contain every possible circumstance. Such analytical skills are useful both for solving problems in this text and for applying physics in everyday and professional life.

Problem-Solving Steps

While there is no simple step-by-step method that works for every problem, the following general procedures facilitate problem solving and make it more meaningful. A certain amount of creativity and insight is required as well.

Step 1

Examine the situation to determine which physical principles are involved. It often helps to *draw a simple sketch* at the outset. You will also need to decide which direction is positive and note that on your sketch. Once you have identified the physical principles, it is much easier to find and apply the equations representing those principles. Although finding the correct equation is essential, keep in mind that equations represent physical principles, laws of nature, and relationships among physical quantities. Without a conceptual understanding of a problem, a numerical solution is meaningless.

Step 2

Make a list of what is given or can be inferred from the problem as stated (identify the knowns). Many problems are stated very succinctly and require some inspection to determine what is known. A sketch can also be very useful at this point. Formally identifying the knowns is of particular importance in applying physics to real-world situations. Remember, “stopped” means velocity is zero, and we often can take initial time and position as zero.

Step 3

Identify exactly what needs to be determined in the problem (identify the unknowns). In complex problems, especially, it is not always obvious what needs to be found or in what sequence. Making a list can help.

Step 4

Find an equation or set of equations that can help you solve the problem. Your list of knowns and unknowns can help here. It is easiest if you can find equations that contain only one unknown—that is, all of the other variables are known, so you can easily solve for the unknown. If the equation contains more than one unknown, then an additional equation is needed to solve the problem. In some problems, several unknowns must be determined to get at the one needed most. In such problems it is especially important to keep physical principles in mind to avoid going astray in a sea of equations. You may have to use two (or more) different equations to get the final answer.

Step 5

Substitute the knowns along with their units into the appropriate equation, and obtain numerical solutions complete with units. This step produces the numerical answer; it also provides a check on units that can help you find errors. If the units of the answer are incorrect, then an error has been made. However, be warned that correct units do not guarantee that the numerical part of the answer is also correct.

Step 6

Check the answer to see if it is reasonable: Does it make sense? This final step is extremely important—the goal of physics is to accurately describe nature. To see if the answer is reasonable, check both its magnitude and its sign, in addition to its units. Your judgment will improve as you solve more and more physics problems, and it will become possible for you to make finer and finer judgments regarding whether nature is adequately described by the answer to a problem. This step brings the problem back to its conceptual meaning. If you can judge whether the answer is reasonable, you have a deeper understanding of physics than just being able to mechanically solve a problem.

When solving problems, we often perform these steps in different order, and we also tend to do several steps simultaneously. There is no rigid procedure that will work every time. Creativity and insight grow with experience, and the basics of problem solving become almost automatic. One way to get practice is to work out the text's examples for yourself as you read. Another is to work as many end-of-section problems as possible, starting with the easiest to build confidence and progressing to the more difficult. Once you become involved in physics, you will see it all around you, and you can begin to apply it to situations you encounter outside the classroom, just as is done in many of the applications in this text.

Unreasonable Results

Physics must describe nature accurately. Some problems have results that are unreasonable because one premise is unreasonable or because certain premises are inconsistent with one another. The physical principle applied correctly then produces an unreasonable result. For example, if a person starting a foot race accelerates at 0.40 m/s^2 for 100 s, his final speed will be 40 m/s (about 150 km/h)—clearly unreasonable because the time of 100 s is an unreasonable premise. The physics is correct in a sense, but there is more to describing nature than just manipulating equations correctly. Checking the result of a problem to see if it is reasonable does more than help uncover errors in problem solving—it also builds intuition in judging whether nature is being accurately described.

Use the following strategies to determine whether an answer is reasonable and, if it is not, to determine what is the cause.

Step 1

Solve the problem using strategies as outlined and in the format followed in the worked examples in the text. In the example given in the preceding paragraph, you would identify the givens as the acceleration and time and use the equation below to find the unknown final velocity. That is,

Equation:

$$v = v_0 + at = 0 + (0.40 \text{ m/s}^2)(100 \text{ s}) = 40 \text{ m/s}.$$

Step 2

Check to see if the answer is reasonable. Is it too large or too small, or does it have the wrong sign, improper units, ...? In this case, you may need to convert meters per second into a more familiar unit, such as miles per hour.

Equation:

$$\left(\frac{40 \text{ m}}{\text{s}}\right) \left(\frac{3.28 \text{ ft}}{\text{m}}\right) \left(\frac{1 \text{ mi}}{5280 \text{ ft}}\right) \left(\frac{60 \text{ s}}{\text{min}}\right) \left(\frac{60 \text{ min}}{1 \text{ h}}\right) = 89 \text{ mph}$$

This velocity is about four times greater than a person can run—so it is too large.

Step 3

If the answer is unreasonable, look for what specifically could cause the identified difficulty. In the example of the runner, there are only two assumptions that are suspect. The acceleration could be too great or the time too long. First look at the acceleration and think about what the number means. If someone accelerates at 0.40 m/s^2 , their velocity is increasing by 0.4 m/s each second. Does this seem reasonable? If so, the time must be too long. It is not possible for someone to accelerate at a constant rate of 0.40 m/s^2 for 100 s (almost two minutes).

Section Summary

- *The six basic problem solving steps for physics are:*

Step 1. Examine the situation to determine which physical principles are involved.

Step 2. Make a list of what is given or can be inferred from the problem as stated (identify the knowns).

Step 3. Identify exactly what needs to be determined in the problem (identify the unknowns).

Step 4. Find an equation or set of equations that can help you solve the problem.

Step 5. Substitute the knowns along with their units into the appropriate equation, and obtain numerical solutions complete with units.

Step 6. Check the answer to see if it is reasonable: Does it make sense?

Conceptual Questions

Exercise:

Problem:

What information do you need in order to choose which equation or equations to use to solve a problem? Explain.

Exercise:

Problem:

What is the last thing you should do when solving a problem? Explain.

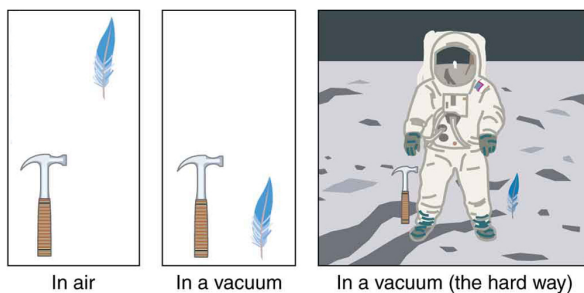
Falling Objects

- Describe the effects of gravity on objects in motion.
- Describe the motion of objects that are in free fall.
- Calculate the position and velocity of objects in free fall.

Falling objects form an interesting class of motion problems. For example, we can estimate the depth of a vertical mine shaft by dropping a rock into it and listening for the rock to hit the bottom. By applying the kinematics developed so far to falling objects, we can examine some interesting situations and learn much about gravity in the process.

Gravity

The most remarkable and unexpected fact about falling objects is that, if air resistance and friction are negligible, then in a given location all objects fall toward the center of Earth with the *same constant acceleration, independent of their mass*. This experimentally determined fact is unexpected, because we are so accustomed to the effects of air resistance and friction that we expect light objects to fall slower than heavy ones.



A hammer and a feather will fall with the same constant acceleration if air resistance is considered negligible. This is a general characteristic of gravity not unique to Earth, as astronaut David R. Scott demonstrated on the Moon in 1971, where the

acceleration due to gravity is
only 1.67 m/s^2 .

In the real world, air resistance can cause a lighter object to fall slower than a heavier object of the same size. A tennis ball will reach the ground after a hard baseball dropped at the same time. (It might be difficult to observe the difference if the height is not large.) Air resistance opposes the motion of an object through the air, while friction between objects—such as between clothes and a laundry chute or between a stone and a pool into which it is dropped—also opposes motion between them. For the ideal situations of these first few chapters, an object *falling without air resistance or friction* is defined to be in **free-fall**.

The force of gravity causes objects to fall toward the center of Earth. The acceleration of free-falling objects is therefore called the **acceleration due to gravity**. The acceleration due to gravity is *constant*, which means we can apply the kinematics equations to any falling object where air resistance and friction are negligible. This opens a broad class of interesting situations to us. The acceleration due to gravity is so important that its magnitude is given its own symbol, g . It is constant at any given location on Earth and has the average value

Equation:

$$g = 9.80 \text{ m/s}^2.$$

Although g varies from 9.78 m/s^2 to 9.83 m/s^2 , depending on latitude, altitude, underlying geological formations, and local topography, the average value of 9.80 m/s^2 will be used in this text unless otherwise specified. The direction of the acceleration due to gravity is *downward (towards the center of Earth)*. In fact, its direction *defines* what we call vertical. Note that whether the acceleration a in the kinematic equations has the value $+g$ or $-g$ depends on how we define our coordinate system. If we define the upward direction as positive, then $a = -g = -9.80 \text{ m/s}^2$, and if we define the downward direction as positive, then $a = g = 9.80 \text{ m/s}^2$.

One-Dimensional Motion Involving Gravity

The best way to see the basic features of motion involving gravity is to start with the simplest situations and then progress toward more complex ones. So we start by considering straight up and down motion with no air resistance or friction. These assumptions mean that the velocity (if there is any) is vertical. If the object is dropped, we know the initial velocity is zero. Once the object has left contact with whatever held or threw it, the object is in free-fall. Under these circumstances, the motion is one-dimensional and has constant acceleration of magnitude g . We will also represent vertical displacement with the symbol y and use x for horizontal displacement.

Note:

Kinematic Equations for Objects in Free-Fall where Acceleration = $-g$

Equation:

$$v = v_0 - gt$$

Equation:

$$y = y_0 + v_0t - \frac{1}{2}gt^2$$

Equation:

$$v^2 = v_0^2 - 2g(y - y_0)$$

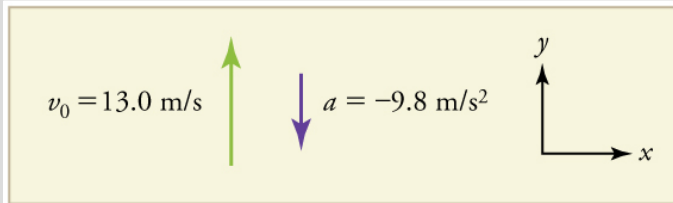
Example:

Calculating Position and Velocity of a Falling Object: A Rock Thrown Upward

A person standing on the edge of a high cliff throws a rock straight up with an initial velocity of 13.0 m/s. The rock misses the edge of the cliff as it falls back to earth. Calculate the position and velocity of the rock 1.00 s, 2.00 s, and 3.00 s after it is thrown, neglecting the effects of air resistance.

Strategy

Draw a sketch.



We are asked to determine the position y at various times. It is reasonable to take the initial position y_0 to be zero. This problem involves one-dimensional motion in the vertical direction. We use plus and minus signs to indicate direction, with up being positive and down negative. Since up is positive, and the rock is thrown upward, the initial velocity must be positive too. The acceleration due to gravity is downward, so a is negative. It is crucial that the initial velocity and the acceleration due to gravity have opposite signs.

Opposite signs indicate that the acceleration due to gravity opposes the initial motion and will slow and eventually reverse it.

Since we are asked for values of position and velocity at three times, we will refer to these as y_1 and v_1 ; y_2 and v_2 ; and y_3 and v_3 .

Solution for Position y_1

1. Identify the knowns. We know that $y_0 = 0$; $v_0 = 13.0 \text{ m/s}$;

$a = -g = -9.80 \text{ m/s}^2$; and $t = 1.00 \text{ s}$.

2. Identify the best equation to use. We will use $y = y_0 + v_0 t + \frac{1}{2} a t^2$ because it includes only one unknown, y (or y_1 , here), which is the value we want to find.

3. Plug in the known values and solve for y_1 .

Equation:

$$y_1 = 0 + (13.0 \text{ m/s})(1.00 \text{ s}) + \frac{1}{2} (-9.80 \text{ m/s}^2)(1.00 \text{ s})^2 = 8.10 \text{ m}$$

Discussion

The rock is 8.10 m above its starting point at $t = 1.00 \text{ s}$, since $y_1 > y_0$. It could be *moving* up or down; the only way to tell is to calculate v_1 and find out if it is positive or negative.

Solution for Velocity v_1

1. Identify the knowns. We know that $y_0 = 0$; $v_0 = 13.0 \text{ m/s}$;

$a = -g = -9.80 \text{ m/s}^2$; and $t = 1.00 \text{ s}$. We also know from the solution above that $y_1 = 8.10 \text{ m}$.

2. Identify the best equation to use. The most straightforward is $v = v_0 - gt$ (from $v = v_0 + at$, where $a = \text{gravitational acceleration} = -g$).
3. Plug in the knowns and solve.

Equation:

$$v_1 = v_0 - gt = 13.0 \text{ m/s} - (9.80 \text{ m/s}^2)(1.00 \text{ s}) = 3.20 \text{ m/s}$$

Discussion

The positive value for v_1 means that the rock is still heading upward at $t = 1.00 \text{ s}$. However, it has slowed from its original 13.0 m/s , as expected.

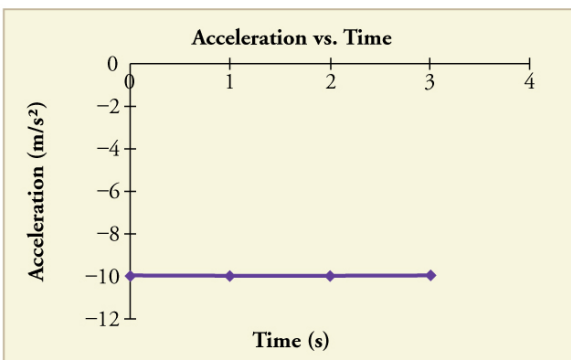
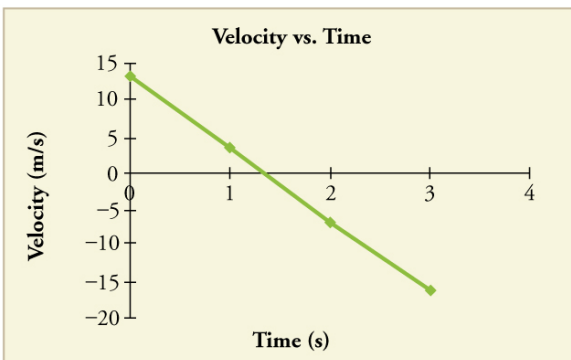
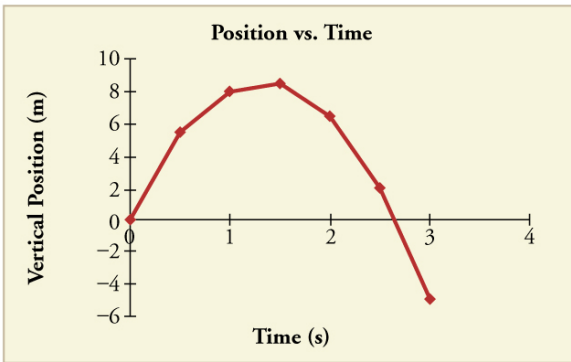
Solution for Remaining Times

The procedures for calculating the position and velocity at $t = 2.00 \text{ s}$ and 3.00 s are the same as those above. The results are summarized in [\[link\]](#) and illustrated in [\[link\]](#).

Time, t	Position, y	Velocity, v	Acceleration, a
1.00 s	8.10 m	3.20 m/s	-9.80 m/s^2
2.00 s	6.40 m	-6.60 m/s	-9.80 m/s^2
3.00 s	-5.10 m	-16.4 m/s	-9.80 m/s^2

Results

Graphing the data helps us understand it more clearly.



Vertical position, vertical velocity, and vertical acceleration vs. time for a rock thrown vertically up at the edge of a cliff. Notice that velocity changes linearly with time and that acceleration is constant.

Misconception Alert! Notice that the position vs. time graph shows vertical position only. It is easy to get the impression that the graph shows some

horizontal motion—the shape of the graph looks like the path of a projectile. But this is not the case; the horizontal axis is *time*, not space. The actual path of the rock in space is straight up, and straight down.

Discussion

The interpretation of these results is important. At 1.00 s the rock is above its starting point and heading upward, since y_1 and v_1 are both positive. At 2.00 s, the rock is still above its starting point, but the negative velocity means it is moving downward. At 3.00 s, both y_3 and v_3 are negative, meaning the rock is below its starting point and continuing to move downward. Notice that when the rock is at its highest point (at 1.5 s), its velocity is zero, but its acceleration is still -9.80 m/s^2 . Its acceleration is -9.80 m/s^2 for the whole trip—while it is moving up and while it is moving down. Note that the values for y are the positions (or displacements) of the rock, not the total distances traveled. Finally, note that free-fall applies to upward motion as well as downward. Both have the same acceleration—the acceleration due to gravity, which remains constant the entire time. Astronauts training in the famous Vomit Comet, for example, experience free-fall while arcing up as well as down, as we will discuss in more detail later.

Note:

Making Connections: Take-Home Experiment—Reaction Time

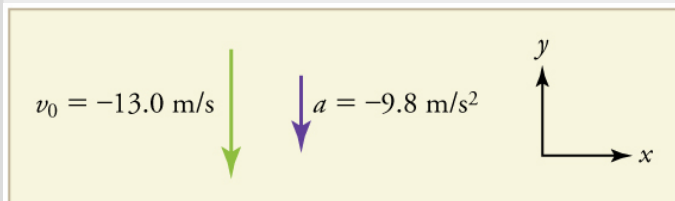
A simple experiment can be done to determine your reaction time. Have a friend hold a ruler between your thumb and index finger, separated by about 1 cm. Note the mark on the ruler that is right between your fingers. Have your friend drop the ruler unexpectedly, and try to catch it between your two fingers. Note the new reading on the ruler. Assuming acceleration is that due to gravity, calculate your reaction time. How far would you travel in a car (moving at 30 m/s) if the time it took your foot to go from the gas pedal to the brake was twice this reaction time?

Example:**Calculating Velocity of a Falling Object: A Rock Thrown Down**

What happens if the person on the cliff throws the rock straight down, instead of straight up? To explore this question, calculate the velocity of the rock when it is 5.10 m below the starting point, and has been thrown downward with an initial speed of 13.0 m/s.

Strategy

Draw a sketch.



Since up is positive, the final position of the rock will be negative because it finishes below the starting point at $y_0 = 0$. Similarly, the initial velocity is downward and therefore negative, as is the acceleration due to gravity. We expect the final velocity to be negative since the rock will continue to move downward.

Solution

1. Identify the knowns. $y_0 = 0$; $y_1 = -5.10$ m; $v_0 = -13.0$ m/s; $a = -g = -9.80$ m/s².
2. Choose the kinematic equation that makes it easiest to solve the problem. The equation $v^2 = v_0^2 + 2a(y - y_0)$ works well because the only unknown in it is v . (We will plug y_1 in for y .)
3. Enter the known values

Equation:

$$v^2 = (-13.0 \text{ m/s})^2 + 2(-9.80 \text{ m/s}^2)(-5.10 \text{ m} - 0 \text{ m}) = 268.96 \text{ m}^2/\text{s}^2,$$

where we have retained extra significant figures because this is an intermediate result.

Taking the square root, and noting that a square root can be positive or negative, gives

Equation:

$$v = \pm 16.4 \text{ m/s}.$$

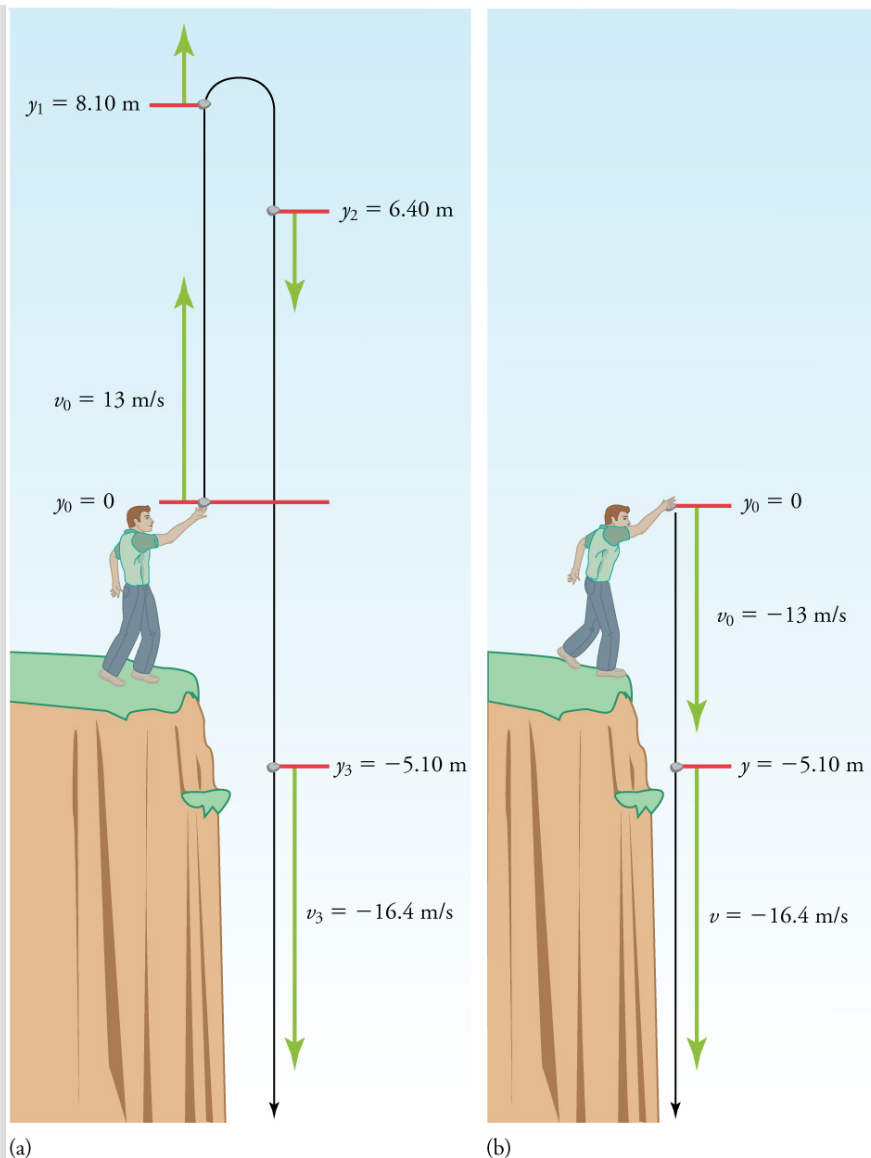
The negative root is chosen to indicate that the rock is still heading down. Thus,

Equation:

$$v = -16.4 \text{ m/s.}$$

Discussion

Note that *this is exactly the same velocity the rock had at this position when it was thrown straight upward with the same initial speed.* (See [\[link\]](#) and [\[link\]](#)(a).) This is not a coincidental result. Because we only consider the acceleration due to gravity in this problem, the *speed* of a falling object depends only on its initial speed and its vertical position relative to the starting point. For example, if the velocity of the rock is calculated at a height of 8.10 m above the starting point (using the method from [\[link\]](#)) when the initial velocity is 13.0 m/s straight up, a result of $\pm 3.20 \text{ m/s}$ is obtained. Here both signs are meaningful; the positive value occurs when the rock is at 8.10 m and heading up, and the negative value occurs when the rock is at 8.10 m and heading back down. It has the same *speed* but the opposite direction.



(a) A person throws a rock straight up, as explored in [\[link\]](#). The arrows are velocity vectors at 0, 1.00, 2.00, and 3.00 s. (b) A person throws a rock straight down from a cliff with the same initial speed as before, as in [\[link\]](#). Note that at the same distance below the point of release, the rock has the same velocity in both cases.

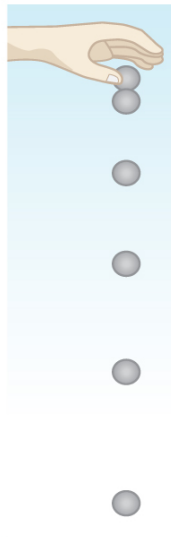
Another way to look at it is this: In [\[link\]](#), the rock is thrown up with an initial velocity of 13.0 m/s . It rises and then falls back down. When its

position is $y = 0$ on its way back down, its velocity is -13.0 m/s . That is, it has the same speed on its way down as on its way up. We would then expect its velocity at a position of $y = -5.10 \text{ m}$ to be the same whether we have thrown it upwards at $+13.0 \text{ m/s}$ or thrown it downwards at -13.0 m/s . The velocity of the rock on its way down from $y = 0$ is the same whether we have thrown it up or down to start with, as long as the speed with which it was initially thrown is the same.

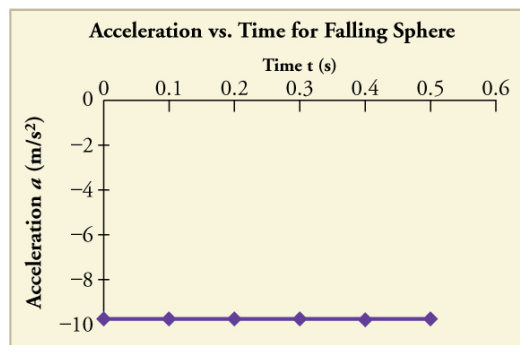
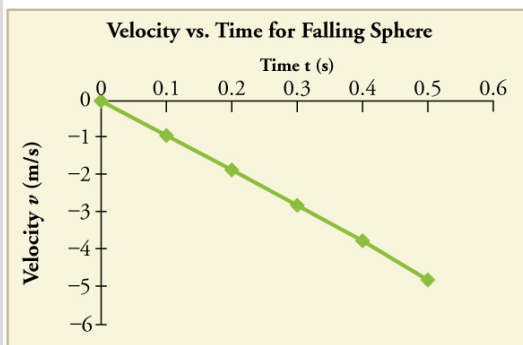
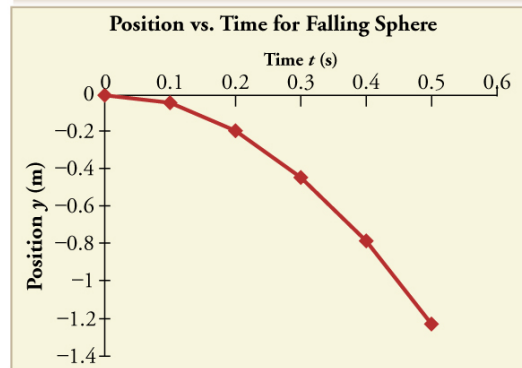
Example:**Find g from Data on a Falling Object**

The acceleration due to gravity on Earth differs slightly from place to place, depending on topography (e.g., whether you are on a hill or in a valley) and subsurface geology (whether there is dense rock like iron ore as opposed to light rock like salt beneath you.) The precise acceleration due to gravity can be calculated from data taken in an introductory physics laboratory course.

An object, usually a metal ball for which air resistance is negligible, is dropped and the time it takes to fall a known distance is measured. See, for example, [\[link\]](#). Very precise results can be produced with this method if sufficient care is taken in measuring the distance fallen and the elapsed time.



y (m)	v (m/s)	t (s)
0	0	0
-0.049	-0.98	0.1
-0.196	-1.96	0.2
-0.441	-2.94	0.3
-0.784	-3.92	0.4
-1.225	-4.90	0.5



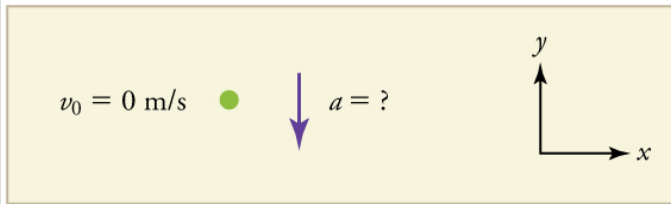
Positions and velocities of a metal ball released from rest when air resistance is negligible. Velocity is seen to increase linearly with time while displacement increases with time squared.

Acceleration is a constant and is equal to gravitational acceleration.

Suppose the ball falls 1.0000 m in 0.45173 s. Assuming the ball is not affected by air resistance, what is the precise acceleration due to gravity at this location?

Strategy

Draw a sketch.



We need to solve for acceleration a . Note that in this case, displacement is downward and therefore negative, as is acceleration.

Solution

1. Identify the knowns. $y_0 = 0$; $y = -1.0000$ m; $t = 0.45173$; $v_0 = 0$.
2. Choose the equation that allows you to solve for a using the known values.

Equation:

$$y = y_0 + v_0 t + \frac{1}{2} a t^2$$

3. Substitute 0 for v_0 and rearrange the equation to solve for a . Substituting 0 for v_0 yields

Equation:

$$y = y_0 + \frac{1}{2} a t^2.$$

Solving for a gives

Equation:

$$a = \frac{2(y - y_0)}{t^2}.$$

4. Substitute known values yields

Equation:

$$a = \frac{2(-1.0000 \text{ m} - 0)}{(0.45173 \text{ s})^2} = -9.8010 \text{ m/s}^2,$$

so, because $a = -g$ with the directions we have chosen,

Equation:

$$g = 9.8010 \text{ m/s}^2.$$

Discussion

The negative value for a indicates that the gravitational acceleration is downward, as expected. We expect the value to be somewhere around the average value of 9.80 m/s^2 , so 9.8010 m/s^2 makes sense. Since the data going into the calculation are relatively precise, this value for g is more precise than the average value of 9.80 m/s^2 ; it represents the local value for the acceleration due to gravity.

Exercise:

Check Your Understanding

Problem:

A chunk of ice breaks off a glacier and falls 30.0 meters before it hits the water. Assuming it falls freely (there is no air resistance), how long does it take to hit the water?

Solution:

We know that initial position $y_0 = 0$, final position $y = -30.0 \text{ m}$, and $a = -g = -9.80 \text{ m/s}^2$. We can then use the equation

$y = y_0 + v_0t + \frac{1}{2}at^2$ to solve for t . Inserting $a = -g$, we obtain

Equation:

$$y = 0 + 0 - \frac{1}{2}gt^2$$

$$t^2 = \frac{2y}{-g}$$

$$t = \pm \sqrt{\frac{2y}{-g}} = \pm \sqrt{\frac{2(-30.0 \text{ m})}{-9.80 \text{ m/s}^2}} = \pm \sqrt{6.12 \text{ s}^2} = 2.47 \text{ s} \approx 2.5 \text{ s}$$

where we take the positive value as the physically relevant answer. Thus, it takes about 2.5 seconds for the piece of ice to hit the water.

Note:

PhET Explorations: Equation Grapher

Learn about graphing polynomials. The shape of the curve changes as the constants are adjusted. View the curves for the individual terms (e.g. $y = bx$) to see how they add to generate the polynomial curve.

https://phet.colorado.edu/sims/equation-grapher/equation-grapher_en.html

Section Summary

- An object in free-fall experiences constant acceleration if air resistance is negligible.
- On Earth, all free-falling objects have an acceleration due to gravity g , which averages

Equation:

$$g = 9.80 \text{ m/s}^2.$$

- Whether the acceleration a should be taken as $+g$ or $-g$ is determined by your choice of coordinate system. If you choose the upward direction as positive, $a = -g = -9.80 \text{ m/s}^2$ is negative. In the opposite case, $a = +g = 9.80 \text{ m/s}^2$ is positive. Since acceleration is constant, the kinematic equations above can be applied with the appropriate $+g$ or $-g$ substituted for a .
- For objects in free-fall, up is normally taken as positive for displacement, velocity, and acceleration.

Conceptual Questions

Exercise:

Problem:

What is the acceleration of a rock thrown straight upward on the way up? At the top of its flight? On the way down?

Exercise:**Problem:**

An object that is thrown straight up falls back to Earth. This is one-dimensional motion. (a) When is its velocity zero? (b) Does its velocity change direction? (c) Does the acceleration due to gravity have the same sign on the way up as on the way down?

Exercise:**Problem:**

Suppose you throw a rock nearly straight up at a coconut in a palm tree, and the rock misses on the way up but hits the coconut on the way down. Neglecting air resistance, how does the speed of the rock when it hits the coconut on the way down compare with what it would have been if it had hit the coconut on the way up? Is it more likely to dislodge the coconut on the way up or down? Explain.

Exercise:**Problem:**

If an object is thrown straight up and air resistance is negligible, then its speed when it returns to the starting point is the same as when it was released. If air resistance were not negligible, how would its speed upon return compare with its initial speed? How would the maximum height to which it rises be affected?

Exercise:**Problem:**

The severity of a fall depends on your speed when you strike the ground. All factors but the acceleration due to gravity being the same, how many times higher could a safe fall on the Moon be than on Earth (gravitational acceleration on the Moon is about $1/6$ that of the Earth)?

Exercise:**Problem:**

How many times higher could an astronaut jump on the Moon than on Earth if his takeoff speed is the same in both locations (gravitational acceleration on the Moon is about $1/6$ of g on Earth)?

Problems & Exercises

Assume air resistance is negligible unless otherwise stated.

Exercise:**Problem:**

Calculate the displacement and velocity at times of (a) 0.500, (b) 1.00, (c) 1.50, and (d) 2.00 s for a ball thrown straight up with an initial velocity of 15.0 m/s. Take the point of release to be $y_0 = 0$.

Solution:

(a) $y_1 = 6.28 \text{ m}$; $v_1 = 10.1 \text{ m/s}$

(b) $y_2 = 10.1 \text{ m}$; $v_2 = 5.20 \text{ m/s}$

(c) $y_3 = 11.5 \text{ m}$; $v_3 = 0.300 \text{ m/s}$

(d) $y_4 = 10.4 \text{ m}$; $v_4 = -4.60 \text{ m/s}$

Exercise:**Problem:**

Calculate the displacement and velocity at times of (a) 0.500, (b) 1.00, (c) 1.50, (d) 2.00, and (e) 2.50 s for a rock thrown straight down with an initial velocity of 14.0 m/s from the Verrazano Narrows Bridge in New York City. The roadway of this bridge is 70.0 m above the water.

Exercise:

Problem:

A basketball referee tosses the ball straight up for the starting tip-off. At what velocity must a basketball player leave the ground to rise 1.25 m above the floor in an attempt to get the ball?

Solution:

$$v_0 = 4.95 \text{ m/s}$$

Exercise:**Problem:**

A rescue helicopter is hovering over a person whose boat has sunk. One of the rescuers throws a life preserver straight down to the victim with an initial velocity of 1.40 m/s and observes that it takes 1.8 s to reach the water. (a) List the knowns in this problem. (b) How high above the water was the preserver released? Note that the downdraft of the helicopter reduces the effects of air resistance on the falling life preserver, so that an acceleration equal to that of gravity is reasonable.

Exercise:**Problem:**

A dolphin in an aquatic show jumps straight up out of the water at a velocity of 13.0 m/s. (a) List the knowns in this problem. (b) How high does his body rise above the water? To solve this part, first note that the final velocity is now a known and identify its value. Then identify the unknown, and discuss how you chose the appropriate equation to solve for it. After choosing the equation, show your steps in solving for the unknown, checking units, and discuss whether the answer is reasonable. (c) How long is the dolphin in the air? Neglect any effects due to his size or orientation.

Solution:

$$(a) a = -9.80 \text{ m/s}^2; v_0 = 13.0 \text{ m/s}; y_0 = 0 \text{ m}$$

(b) $v = 0\text{ m/s}$. Unknown is distance y to top of trajectory, where velocity is zero. Use equation $v^2 = v_0^2 + 2a(y - y_0)$ because it contains all known values except for y , so we can solve for y . Solving for y gives

Equation:

$$v^2 - v_0^2 = 2a(y - y_0)$$

$$\frac{v^2 - v_0^2}{2a} = y - y_0$$

$$y = y_0 + \frac{v^2 - v_0^2}{2a} = 0\text{ m} + \frac{(0\text{ m/s})^2 - (13.0\text{ m/s})^2}{2(-9.80\text{ m/s}^2)} = 8.62\text{ m}$$

Dolphins measure about 2 meters long and can jump several times their length out of the water, so this is a reasonable result.

(c) 2.65 s

Exercise:

Problem:

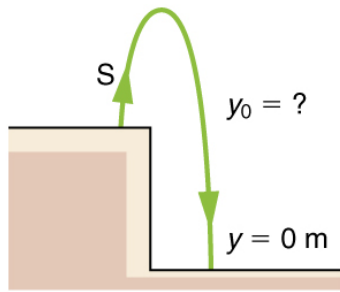
A swimmer bounces straight up from a diving board and falls feet first into a pool. She starts with a velocity of 4.00 m/s, and her takeoff point is 1.80 m above the pool. (a) How long are her feet in the air? (b) What is her highest point above the board? (c) What is her velocity when her feet hit the water?

Exercise:

Problem:

(a) Calculate the height of a cliff if it takes 2.35 s for a rock to hit the ground when it is thrown straight up from the cliff with an initial velocity of 8.00 m/s. (b) How long would it take to reach the ground if it is thrown straight down with the same speed?

Solution:



(a) 8.26 m

(b) 0.717 s

Exercise:

Problem:

A very strong, but inept, shot putter puts the shot straight up vertically with an initial velocity of 11.0 m/s. How long does he have to get out of the way if the shot was released at a height of 2.20 m, and he is 1.80 m tall?

Exercise:

Problem:

You throw a ball straight up with an initial velocity of 15.0 m/s. It passes a tree branch on the way up at a height of 7.00 m. How much additional time will pass before the ball passes the tree branch on the way back down?

Solution:

1.91 s

Exercise:

Problem:

A kangaroo can jump over an object 2.50 m high. (a) Calculate its vertical speed when it leaves the ground. (b) How long is it in the air?

Exercise:

Problem:

Standing at the base of one of the cliffs of Mt. Arapiles in Victoria, Australia, a hiker hears a rock break loose from a height of 105 m. He can't see the rock right away but then does, 1.50 s later. (a) How far above the hiker is the rock when he can see it? (b) How much time does he have to move before the rock hits his head?

Solution:

(a) 94.0 m

(b) 3.13 s

Exercise:**Problem:**

An object is dropped from a height of 75.0 m above ground level. (a) Determine the distance traveled during the first second. (b) Determine the final velocity at which the object hits the ground. (c) Determine the distance traveled during the last second of motion before hitting the ground.

Exercise:**Problem:**

There is a 250-m-high cliff at Half Dome in Yosemite National Park in California. Suppose a boulder breaks loose from the top of this cliff. (a) How fast will it be going when it strikes the ground? (b) Assuming a reaction time of 0.300 s, how long will a tourist at the bottom have to get out of the way after hearing the sound of the rock breaking loose (neglecting the height of the tourist, which would become negligible anyway if hit)? The speed of sound is 335 m/s on this day.

Solution:

(a) -70.0 m/s (downward)

(b) 6.10 s

Exercise:**Problem:**

A ball is thrown straight up. It passes a 2.00-m-high window 7.50 m off the ground on its path up and takes 0.312 s to go past the window. What was the ball's initial velocity? Hint: First consider only the distance along the window, and solve for the ball's velocity at the bottom of the window. Next, consider only the distance from the ground to the bottom of the window, and solve for the initial velocity using the velocity at the bottom of the window as the final velocity.

Exercise:**Problem:**

Suppose you drop a rock into a dark well and, using precision equipment, you measure the time for the sound of a splash to return. (a) Neglecting the time required for sound to travel up the well, calculate the distance to the water if the sound returns in 2.0000 s. (b) Now calculate the distance taking into account the time for sound to travel up the well. The speed of sound is 332.00 m/s in this well.

Solution:

(a) 19.6 m

(b) 18.5 m

Exercise:**Problem:**

A steel ball is dropped onto a hard floor from a height of 1.50 m and rebounds to a height of 1.45 m. (a) Calculate its velocity just before it strikes the floor. (b) Calculate its velocity just after it leaves the floor on its way back up. (c) Calculate its acceleration during contact with the floor if that contact lasts 0.0800 ms (8.00×10^{-5} s). (d) How much did the ball compress during its collision with the floor, assuming the floor is absolutely rigid?

Exercise:

Problem:

A coin is dropped from a hot-air balloon that is 300 m above the ground and rising at 10.0 m/s upward. For the coin, find (a) the maximum height reached, (b) its position and velocity 4.00 s after being released, and (c) the time before it hits the ground.

Solution:

(a) 305 m

(b) 262 m, -29.2 m/s

(c) 8.91 s

Exercise:**Problem:**

A soft tennis ball is dropped onto a hard floor from a height of 1.50 m and rebounds to a height of 1.10 m. (a) Calculate its velocity just before it strikes the floor. (b) Calculate its velocity just after it leaves the floor on its way back up. (c) Calculate its acceleration during contact with the floor if that contact lasts 3.50 ms (3.50×10^{-3} s). (d) How much did the ball compress during its collision with the floor, assuming the floor is absolutely rigid?

Glossary

free-fall

the state of movement that results from gravitational force only

acceleration due to gravity

acceleration of an object as a result of gravity

Graphical Analysis of One-Dimensional Motion

- Describe a straight-line graph in terms of its slope and y-intercept.
- Determine average velocity or instantaneous velocity from a graph of position vs. time.
- Determine average or instantaneous acceleration from a graph of velocity vs. time.
- Derive a graph of velocity vs. time from a graph of position vs. time.
- Derive a graph of acceleration vs. time from a graph of velocity vs. time.

A graph, like a picture, is worth a thousand words. Graphs not only contain numerical information; they also reveal relationships between physical quantities. This section uses graphs of position, velocity, and acceleration versus time to illustrate one-dimensional kinematics.

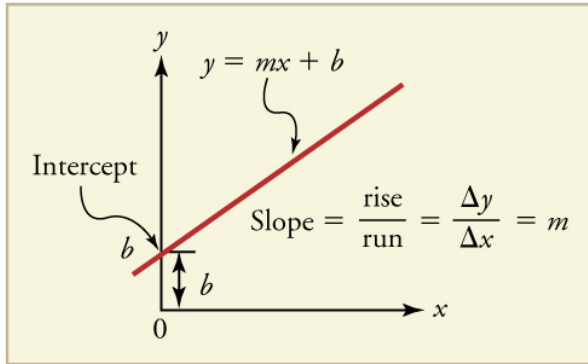
Slopes and General Relationships

First note that graphs in this text have perpendicular axes, one horizontal and the other vertical. When two physical quantities are plotted against one another in such a graph, the horizontal axis is usually considered to be an **independent variable** and the vertical axis a **dependent variable**. If we call the horizontal axis the x -axis and the vertical axis the y -axis, as in [\[link\]](#), a straight-line graph has the general form

Equation:

$$y = mx + b.$$

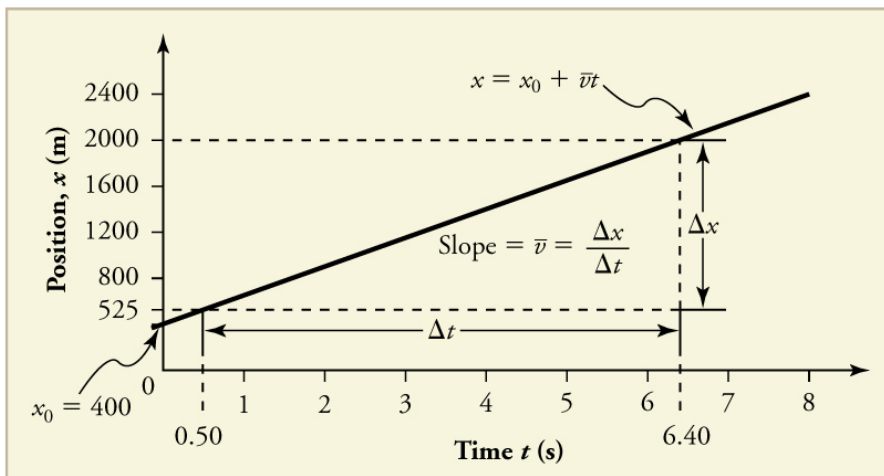
Here m is the **slope**, defined to be the rise divided by the run (as seen in the figure) of the straight line. The letter b is used for the **y-intercept**, which is the point at which the line crosses the vertical axis.



A straight-line graph. The equation for a straight line is $y = mx + b$.

Graph of Position vs. Time ($a = 0$, so v is constant)

Time is usually an independent variable that other quantities, such as position, depend upon. A graph of position versus time would, thus, have x on the vertical axis and t on the horizontal axis. [\[link\]](#) is just such a straight-line graph. It shows a graph of position versus time for a jet-powered car on a very flat dry lake bed in Nevada.



Graph of position versus time for a jet-powered car on the Bonneville Salt Flats.

Using the relationship between dependent and independent variables, we see that the slope in the graph above is average velocity \bar{v} and the intercept is position at time zero—that is, x_0 . Substituting these symbols into $y = mx + b$ gives

Equation:

$$x = \bar{v}t + x_0$$

or

Equation:

$$x = x_0 + \bar{v}t.$$

Thus a graph of position versus time gives a general relationship among displacement(change in position), velocity, and time, as well as giving detailed numerical information about a specific situation.

Note:

The Slope of x vs. t

The slope of the graph of position x vs. time t is velocity v .

Equation:

$$\text{slope} = \frac{\Delta x}{\Delta t} = v$$

Notice that this equation is the same as that derived algebraically from other motion equations in [Motion Equations for Constant Acceleration in One Dimension](#).

From the figure we can see that the car has a position of 25 m at 0.50 s and 2000 m at 6.40 s. Its position at other times can be read from the graph; furthermore, information about its velocity and acceleration can also be obtained from the graph.

Example:

**Determining Average Velocity from a Graph of Position versus Time:
Jet Car**

Find the average velocity of the car whose position is graphed in [\[link\]](#).

Strategy

The slope of a graph of x vs. t is average velocity, since slope equals rise over run. In this case, rise = change in position and run = change in time, so that

Equation:

$$\text{slope} = \frac{\Delta x}{\Delta t} = \bar{v}.$$

Since the slope is constant here, any two points on the graph can be used to find the slope. (Generally speaking, it is most accurate to use two widely separated points on the straight line. This is because any error in reading data from the graph is proportionally smaller if the interval is larger.)

Solution

1. Choose two points on the line. In this case, we choose the points labeled on the graph: (6.4 s, 2000 m) and (0.50 s, 525 m). (Note, however, that you could choose any two points.)
2. Substitute the x and t values of the chosen points into the equation. Remember in calculating change (Δ) we always use final value minus initial value.

Equation:

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{2000 \text{ m} - 525 \text{ m}}{6.4 \text{ s} - 0.50 \text{ s}},$$

yielding

Equation:

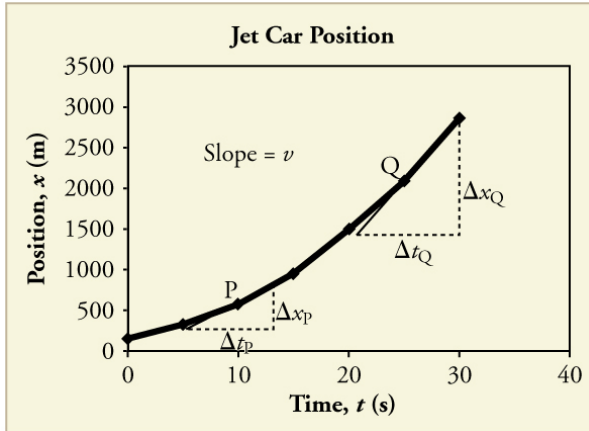
$$\bar{v} = 250 \text{ m/s.}$$

Discussion

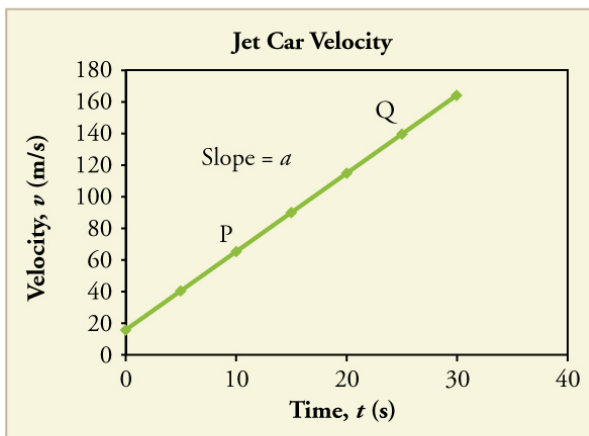
This is an impressively large land speed (900 km/h, or about 560 mi/h): much greater than the typical highway speed limit of 60 mi/h (27 m/s or 96 km/h), but considerably shy of the record of 343 m/s (1234 km/h or 766 mi/h) set in 1997.

Graphs of Motion when a is constant but $a \neq 0$

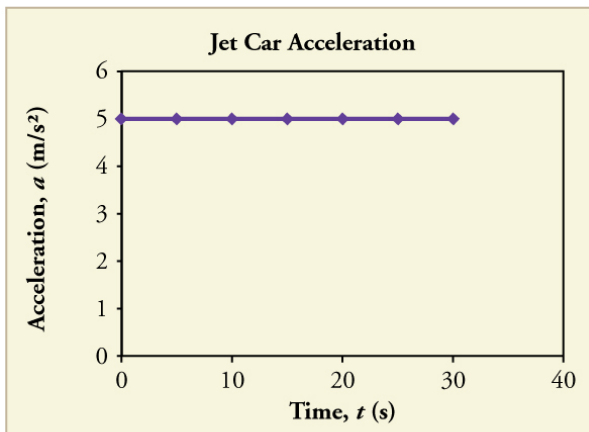
The graphs in [\[link\]](#) below represent the motion of the jet-powered car as it accelerates toward its top speed, but only during the time when its acceleration is constant. Time starts at zero for this motion (as if measured with a stopwatch), and the position and velocity are initially 200 m and 15 m/s, respectively.



(a)



(b)



(c)

Graphs of motion of a jet-powered car during the time span when its acceleration is constant. (a) The slope of an x vs. t graph is velocity. This is

shown at two points, and the instantaneous velocities obtained are plotted in the next graph. Instantaneous velocity at any point is the slope of the tangent at that point. (b) The slope of the v vs. t graph is constant for this part of the motion, indicating constant acceleration. (c) Acceleration has the constant value of 5.0 m/s^2 over the time interval plotted.



A U.S. Air Force jet car speeds down a track. (credit: Matt Trostle, Flickr)

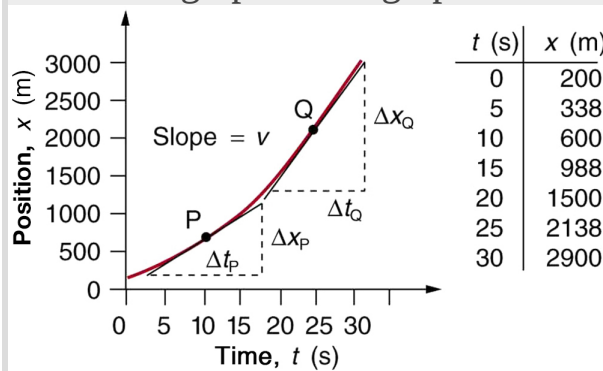
The graph of position versus time in [\[link\]](#)(a) is a curve rather than a straight line. The slope of the curve becomes steeper as time progresses,

showing that the velocity is increasing over time. The slope at any point on a position-versus-time graph is the instantaneous velocity at that point. It is found by drawing a straight line tangent to the curve at the point of interest and taking the slope of this straight line. Tangent lines are shown for two points in [\[link\]\(a\)](#). If this is done at every point on the curve and the values are plotted against time, then the graph of velocity versus time shown in [\[link\]\(b\)](#) is obtained. Furthermore, the slope of the graph of velocity versus time is acceleration, which is shown in [\[link\]\(c\)](#).

Example:

Determining Instantaneous Velocity from the Slope at a Point: Jet Car

Calculate the velocity of the jet car at a time of 25 s by finding the slope of the x vs. t graph in the graph below.



The slope of an x vs. t graph is velocity. This is shown at two points. Instantaneous velocity at any point is the slope of the tangent at that point.

Strategy

The slope of a curve at a point is equal to the slope of a straight line tangent to the curve at that point. This principle is illustrated in [\[link\]](#), where Q is the point at $t = 25$ s.

Solution

1. Find the tangent line to the curve at $t = 25$ s.

2. Determine the endpoints of the tangent. These correspond to a position of 1300 m at time 19 s and a position of 3120 m at time 32 s.
3. Plug these endpoints into the equation to solve for the slope, v .

Equation:

$$\text{slope} = v_Q = \frac{\Delta x_Q}{\Delta t_Q} = \frac{(3120 \text{ m} - 1300 \text{ m})}{(32 \text{ s} - 19 \text{ s})}$$

Thus,

Equation:

$$v_Q = \frac{1820 \text{ m}}{13 \text{ s}} = 140 \text{ m/s}.$$

Discussion

This is the value given in this figure's table for v at $t = 25$ s. The value of 140 m/s for v_Q is plotted in [\[link\]](#). The entire graph of v vs. t can be obtained in this fashion.

Carrying this one step further, we note that the slope of a velocity versus time graph is acceleration. Slope is rise divided by run; on a v vs. t graph, rise = change in velocity Δv and run = change in time Δt .

Note:

The Slope of v vs. t

The slope of a graph of velocity v vs. time t is acceleration a .

Equation:

$$\text{slope} = \frac{\Delta v}{\Delta t} = a$$

Since the velocity versus time graph in [\[link\]](#)(b) is a straight line, its slope is the same everywhere, implying that acceleration is constant. Acceleration versus time is graphed in [\[link\]](#)(c).

Additional general information can be obtained from [\[link\]](#) and the expression for a straight line, $y = mx + b$.

In this case, the vertical axis y is V , the intercept b is v_0 , the slope m is a , and the horizontal axis x is t . Substituting these symbols yields

Equation:

$$v = v_0 + at.$$

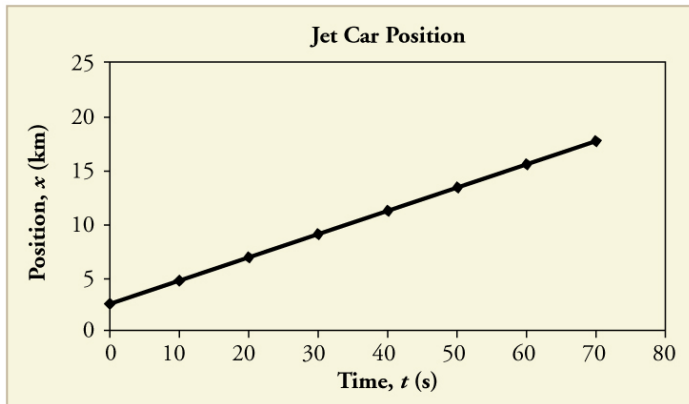
A general relationship for velocity, acceleration, and time has again been obtained from a graph. Notice that this equation was also derived algebraically from other motion equations in [Motion Equations for Constant Acceleration in One Dimension](#).

It is not accidental that the same equations are obtained by graphical analysis as by algebraic techniques. In fact, an important way to *discover* physical relationships is to measure various physical quantities and then make graphs of one quantity against another to see if they are correlated in any way. Correlations imply physical relationships and might be shown by smooth graphs such as those above. From such graphs, mathematical relationships can sometimes be postulated. Further experiments are then performed to determine the validity of the hypothesized relationships.

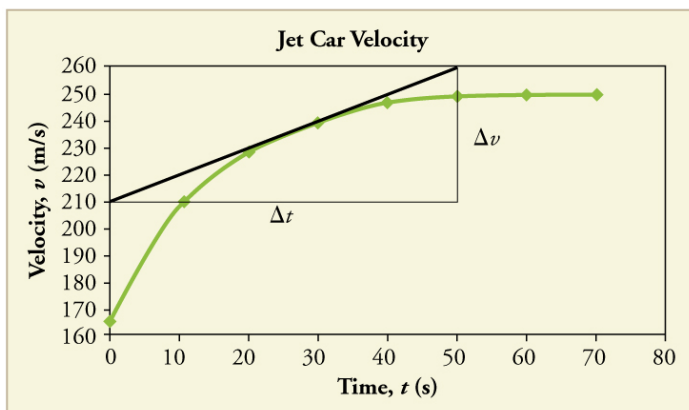
Graphs of Motion Where Acceleration is Not Constant

Now consider the motion of the jet car as it goes from 165 m/s to its top velocity of 250 m/s, graphed in [\[link\]](#). Time again starts at zero, and the initial position and velocity are 2900 m and 165 m/s, respectively. (These were the final position and velocity of the car in the motion graphed in [\[link\]](#).) Acceleration gradually decreases from 5.0 m/s^2 to zero when the car hits 250 m/s. The slope of the x vs. t graph increases until $t = 55 \text{ s}$, after which time the slope is constant. Similarly, velocity increases until 55

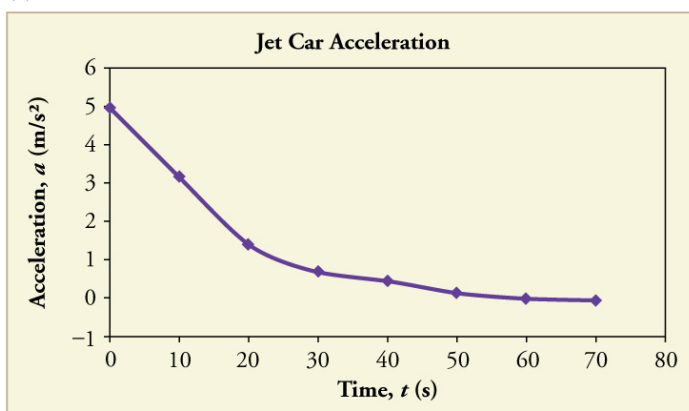
s and then becomes constant, since acceleration decreases to zero at 55 s and remains zero afterward.



(a)



(b)



(c)

Graphs of motion of a jet-powered car as it reaches its top velocity. This motion begins where the motion in

[\[link\]](#) ends. (a) The slope of this graph is velocity; it is plotted in the next graph. (b) The velocity gradually approaches its top value. The slope of this graph is acceleration; it is plotted in the final graph. (c) Acceleration gradually declines to zero when velocity becomes constant.

Example:**Calculating Acceleration from a Graph of Velocity versus Time**

Calculate the acceleration of the jet car at a time of 25 s by finding the slope of the v vs. t graph in [\[link\]](#)(b).

Strategy

The slope of the curve at $t = 25$ s is equal to the slope of the line tangent at that point, as illustrated in [\[link\]](#)(b).

Solution

Determine endpoints of the tangent line from the figure, and then plug them into the equation to solve for slope, a .

Equation:

$$\text{slope} = \frac{\Delta v}{\Delta t} = \frac{(260 \text{ m/s} - 210 \text{ m/s})}{(51 \text{ s} - 1.0 \text{ s})}$$

Equation:

$$a = \frac{50 \text{ m/s}}{50 \text{ s}} = 1.0 \text{ m/s}^2.$$

Discussion

Note that this value for a is consistent with the value plotted in [\[link\]](#)(c) at $t = 25$ s.

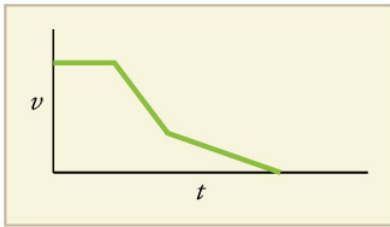
A graph of position versus time can be used to generate a graph of velocity versus time, and a graph of velocity versus time can be used to generate a graph of acceleration versus time. We do this by finding the slope of the graphs at every point. If the graph is linear (i.e., a line with a constant slope), it is easy to find the slope at any point and you have the slope for every point. Graphical analysis of motion can be used to describe both specific and general characteristics of kinematics. Graphs can also be used for other topics in physics. An important aspect of exploring physical relationships is to graph them and look for underlying relationships.

Exercise:

Check Your Understanding

Problem:

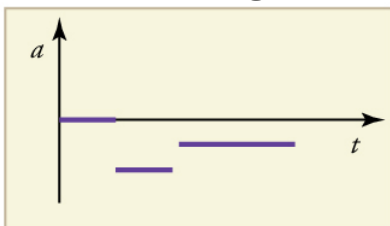
A graph of velocity vs. time of a ship coming into a harbor is shown below. (a) Describe the motion of the ship based on the graph. (b) What would a graph of the ship's acceleration look like?



Solution:

(a) The ship moves at constant velocity and then begins to decelerate at a constant rate. At some point, its deceleration rate decreases. It maintains this lower deceleration rate until it stops moving.

(b) A graph of acceleration vs. time would show zero acceleration in the first leg, large and constant negative acceleration in the second leg, and constant negative acceleration.



Section Summary

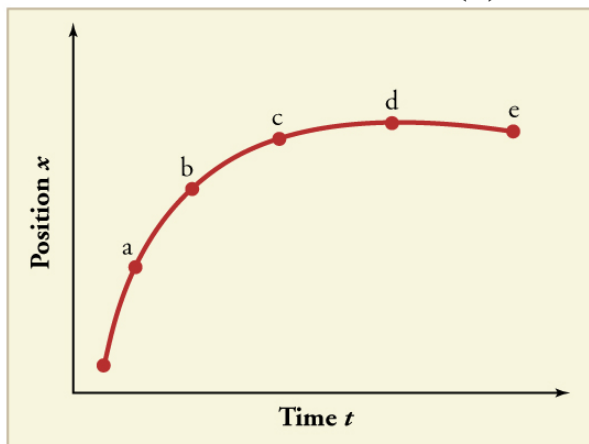
- Graphs of motion can be used to analyze motion.
- Graphical solutions yield identical solutions to mathematical methods for deriving motion equations.
- The slope of a graph of displacement x vs. time t is velocity v .
- The slope of a graph of velocity v vs. time t graph is acceleration a .
- Average velocity, instantaneous velocity, and acceleration can all be obtained by analyzing graphs.

Conceptual Questions

Exercise:

Problem:

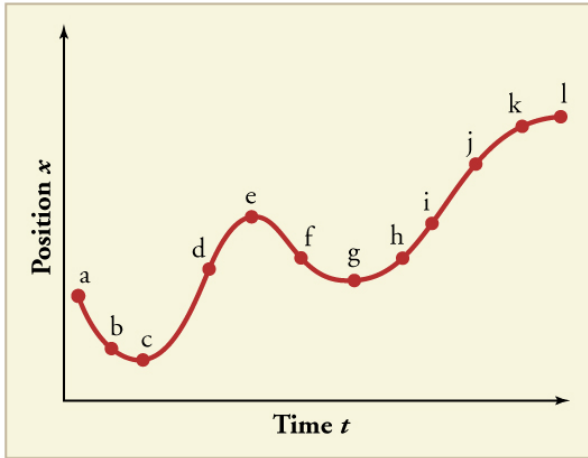
(a) Explain how you can use the graph of position versus time in [\[link\]](#) to describe the change in velocity over time. Identify (b) the time (t_a , t_b , t_c , t_d , or t_e) at which the instantaneous velocity is greatest, (c) the time at which it is zero, and (d) the time at which it is negative.



Exercise:

Problem:

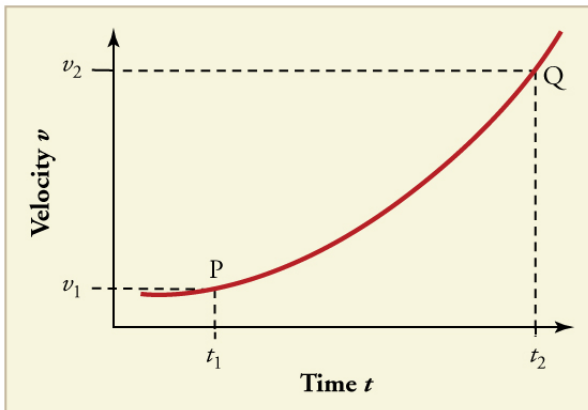
(a) Sketch a graph of velocity versus time corresponding to the graph of position versus time given in [\[link\]](#). (b) Identify the time or times (t_a , t_b , t_c , etc.) at which the instantaneous velocity is greatest. (c) At which times is it zero? (d) At which times is it negative?



Exercise:

Problem:

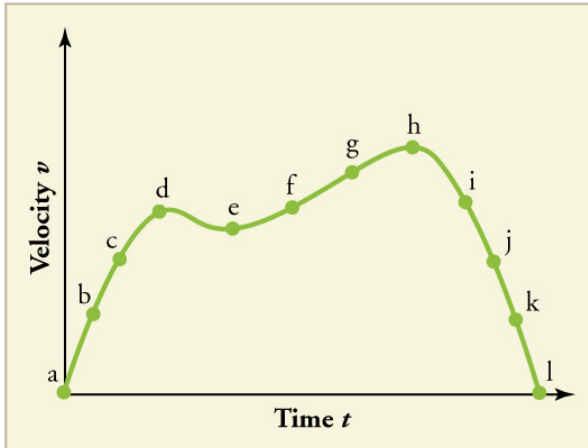
(a) Explain how you can determine the acceleration over time from a velocity versus time graph such as the one in [\[link\]](#). (b) Based on the graph, how does acceleration change over time?



Exercise:

Problem:

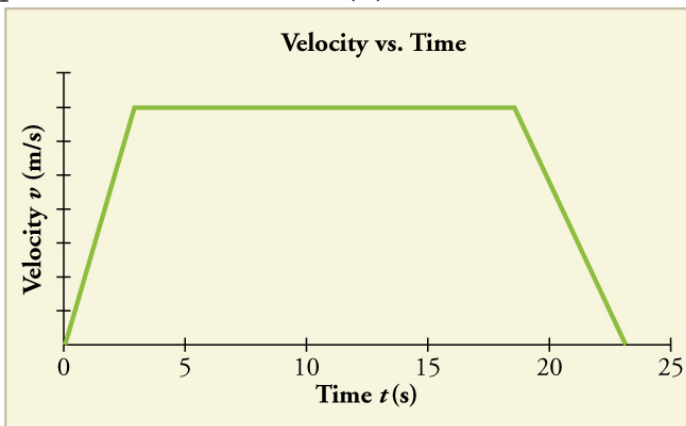
(a) Sketch a graph of acceleration versus time corresponding to the graph of velocity versus time given in [\[link\]](#). (b) Identify the time or times (t_a , t_b , t_c , etc.) at which the acceleration is greatest. (c) At which times is it zero? (d) At which times is it negative?



Exercise:

Problem:

Consider the velocity vs. time graph of a person in an elevator shown in [\[link\]](#). Suppose the elevator is initially at rest. It then accelerates for 3 seconds, maintains that velocity for 15 seconds, then decelerates for 5 seconds until it stops. The acceleration for the entire trip is not constant so we cannot use the equations of motion from [Motion Equations for Constant Acceleration in One Dimension](#) for the complete trip. (We could, however, use them in the three individual sections where acceleration is a constant.) Sketch graphs of (a) position vs. time and (b) acceleration vs. time for this trip.



Exercise:

Problem:

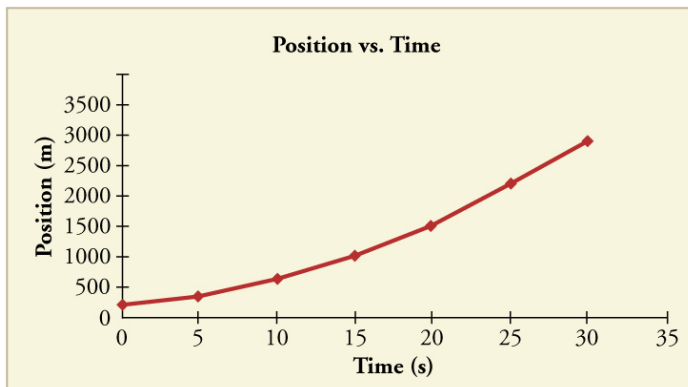
A cylinder is given a push and then rolls up an inclined plane. If the origin is the starting point, sketch the position, velocity, and acceleration of the cylinder vs. time as it goes up and then down the plane.

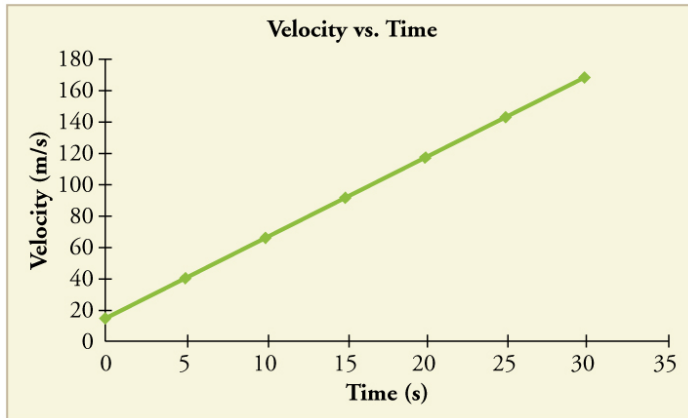
Problems & Exercises

Note: There is always uncertainty in numbers taken from graphs. If your answers differ from expected values, examine them to see if they are within data extraction uncertainties estimated by you.

Exercise:**Problem:**

(a) By taking the slope of the curve in [\[link\]](#), verify that the velocity of the jet car is 115 m/s at $t = 20$ s. (b) By taking the slope of the curve at any point in [\[link\]](#), verify that the jet car's acceleration is 5.0 m/s^2 .





Solution:

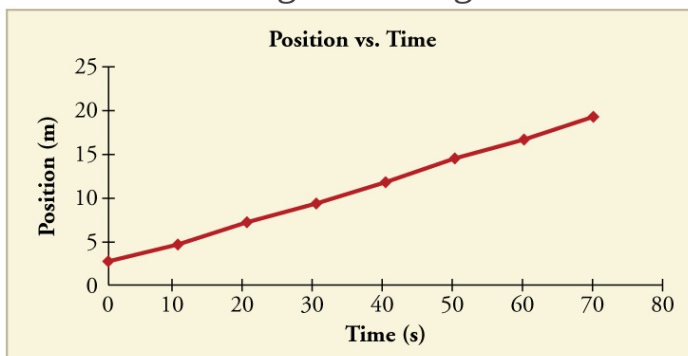
(a) 115 m/s

(b) 5.0 m/s²

Exercise:

Problem:

Using approximate values, calculate the slope of the curve in [\[link\]](#) to verify that the velocity at $t = 10.0$ s is 0.208 m/s. Assume all values are known to 3 significant figures.



Exercise:

Problem:

Using approximate values, calculate the slope of the curve in [\[link\]](#) to verify that the velocity at $t = 30.0$ s is approximately 0.24 m/s.

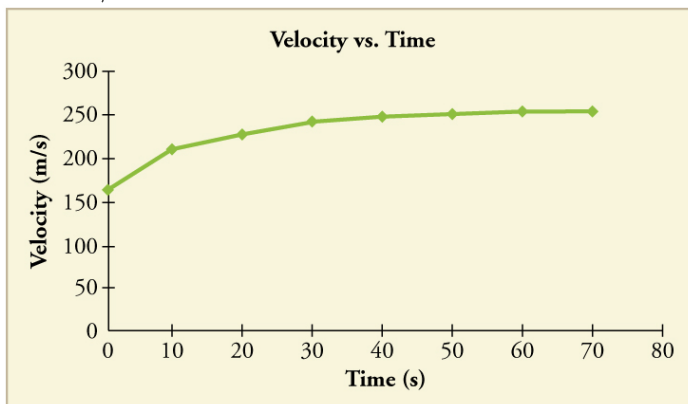
Solution:
Equation:

$$v = \frac{(11.7 - 6.95) \times 10^3 \text{ m}}{(40.0 - 20.0) \text{ s}} = 238 \text{ m/s}$$

Exercise:

Problem:

By taking the slope of the curve in [\[link\]](#), verify that the acceleration is 3.2 m/s^2 at $t = 10 \text{ s}$.

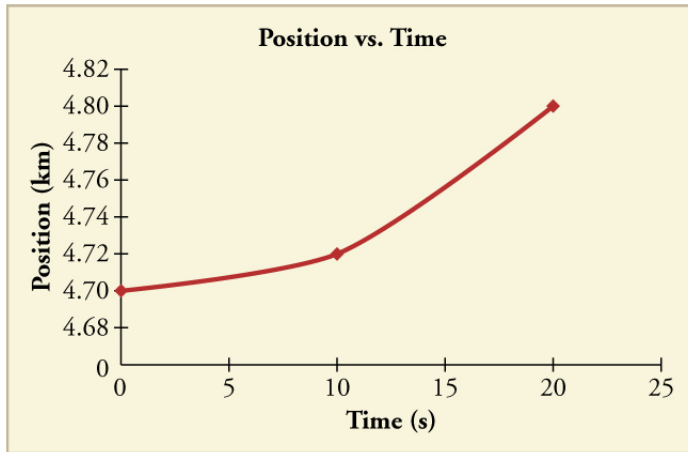


Exercise:

Problem:

Construct the position graph for the subway shuttle train as shown in [\[link\]](#)(a). Your graph should show the position of the train, in kilometers, from $t = 0$ to 20 s . You will need to use the information on acceleration and velocity given in the examples for this figure.

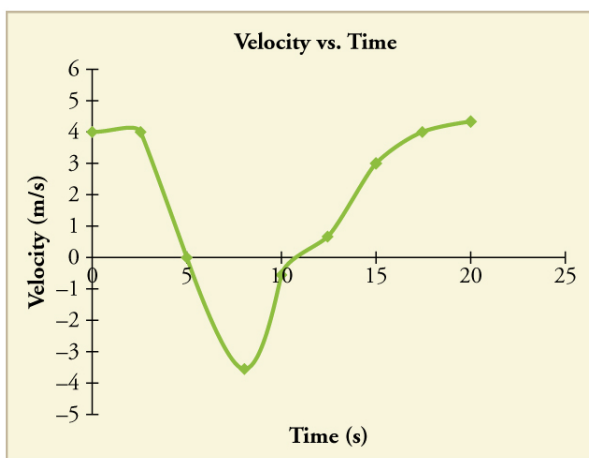
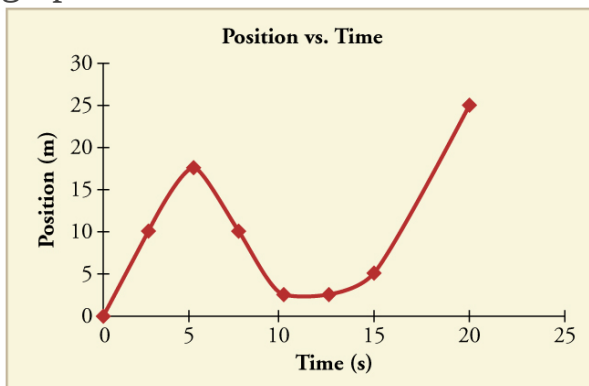
Solution:

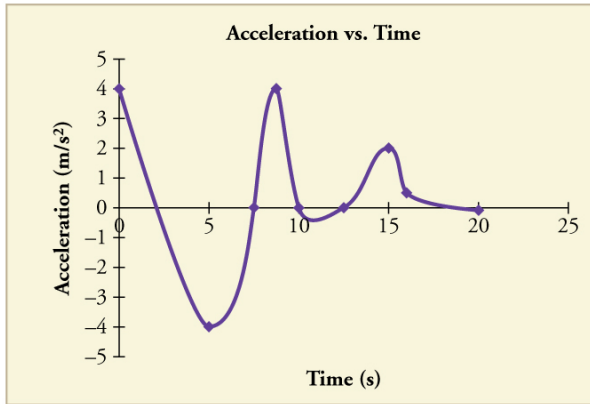


Exercise:

Problem:

(a) Take the slope of the curve in [\[link\]](#) to find the jogger's velocity at $t = 2.5$ s. (b) Repeat at 7.5 s. These values must be consistent with the graph in [\[link\]](#).

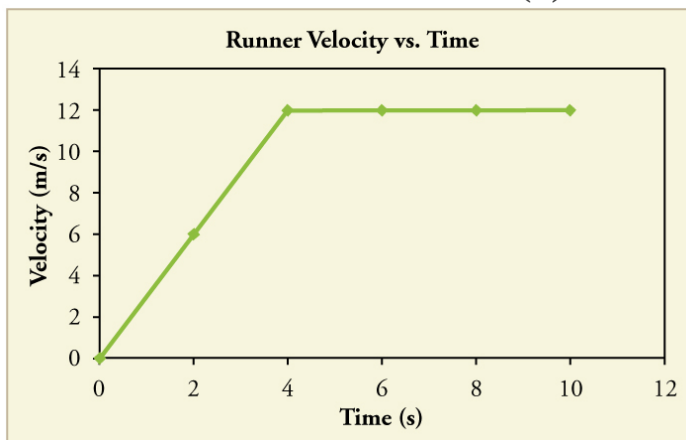




Exercise:

Problem:

A graph of $v(t)$ is shown for a world-class track sprinter in a 100-m race. (See [link](#)). (a) What is his average velocity for the first 4 s? (b) What is his instantaneous velocity at $t = 5$ s? (c) What is his average acceleration between 0 and 4 s? (d) What is his time for the race?



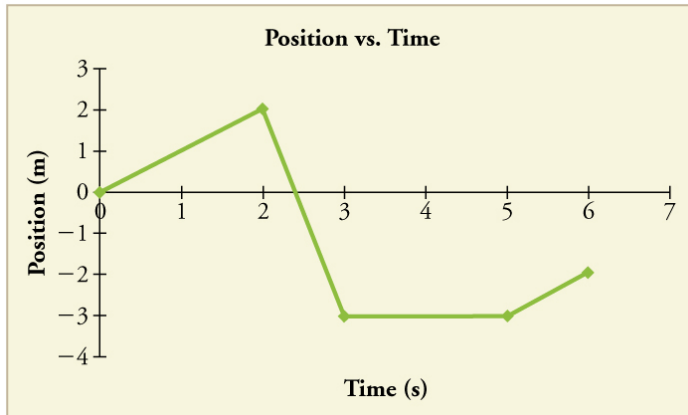
Solution:

- (a) 6 m/s
- (b) 12 m/s
- (c) 3 m/s^2
- (d) 10 s

Exercise:

Problem:

[\[link\]](#) shows the position graph for a particle for 5 s. Draw the corresponding velocity and acceleration graphs.



Glossary

independent variable

the variable that the dependent variable is measured with respect to;
usually plotted along the x -axis

dependent variable

the variable that is being measured; usually plotted along the y -axis

slope

the difference in y -value (the rise) divided by the difference in x -value (the run) of two points on a straight line

y -intercept

the y -value when $x = 0$, or when the graph crosses the y -axis

Introduction to Two-Dimensional Kinematics

class="introduction"

Everyday motion
that we experience
is, thankfully,
rarely as tortuous
as a rollercoaster
ride like this—the
Dragon Khan in
Spain's Universal
Port Aventura
Amusement Park.

However, most
motion is in
curved, rather than
straight-line, paths.

Motion along a
curved path is two-
or three-
dimensional
motion, and can be
described in a
similar fashion to
one-dimensional
motion. (credit:
Boris23/Wikimedi
a Commons)



The arc of a basketball, the orbit of a satellite, a bicycle rounding a curve, a swimmer diving into a pool, blood gushing out of a wound, and a puppy chasing its tail are but a few examples of motions along curved paths. In fact, most motions in nature follow curved paths rather than straight lines. Motion along a curved path on a flat surface or a plane (such as that of a ball on a pool table or a skater on an ice rink) is two-dimensional, and thus described by two-dimensional kinematics. Motion not confined to a plane, such as a car following a winding mountain road, is described by three-dimensional kinematics. Both two- and three-dimensional kinematics are simple extensions of the one-dimensional kinematics developed for straight-line motion in the previous chapter. This simple extension will allow us to apply physics to many more situations, and it will also yield unexpected insights about nature.

Kinematics in Two Dimensions: An Introduction

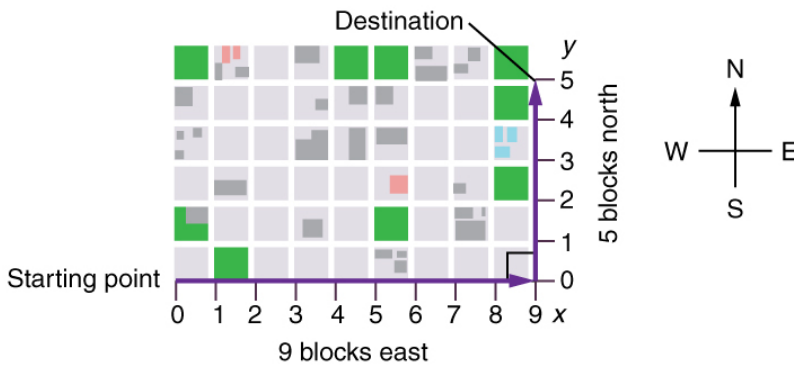
- Observe that motion in two dimensions consists of horizontal and vertical components.
- Understand the independence of horizontal and vertical vectors in two-dimensional motion.



Walkers and drivers in a city like New York are rarely able to travel in straight lines to reach their destinations. Instead, they must follow roads and sidewalks, making two-dimensional, zigzagged paths.
(credit: Margaret W. Carruthers)

Two-Dimensional Motion: Walking in a City

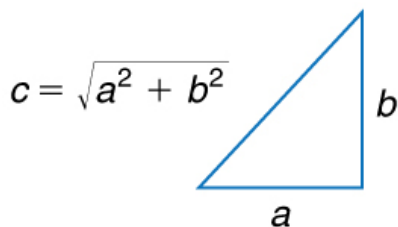
Suppose you want to walk from one point to another in a city with uniform square blocks, as pictured in [\[link\]](#).



A pedestrian walks a two-dimensional path between two points in a city. In this scene, all blocks are square and are the same size.

The straight-line path that a helicopter might fly is blocked to you as a pedestrian, and so you are forced to take a two-dimensional path, such as the one shown. You walk 14 blocks in all, 9 east followed by 5 north. What is the straight-line distance?

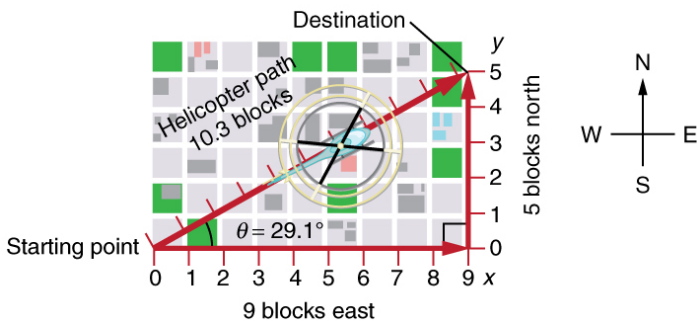
An old adage states that the shortest distance between two points is a straight line. The two legs of the trip and the straight-line path form a right triangle, and so the Pythagorean theorem, $a^2 + b^2 = c^2$, can be used to find the straight-line distance.



The Pythagorean theorem relates the length of the legs of a right triangle,

labeled a and b ,
 with the
 hypotenuse, labeled
 c . The relationship
 is given by:
 $a^2 + b^2 = c^2$. This
 can be rewritten,
 solving for c :
 $c = \sqrt{a^2 + b^2}$.

The hypotenuse of the triangle is the straight-line path, and so in this case its length in units of city blocks is
 $\sqrt{(9 \text{ blocks})^2 + (5 \text{ blocks})^2} = 10.3 \text{ blocks}$, considerably shorter than the 14 blocks you walked. (Note that we are using three significant figures in the answer. Although it appears that “9” and “5” have only one significant digit, they are discrete numbers. In this case “9 blocks” is the same as “9.0 or 9.00 blocks.” We have decided to use three significant figures in the answer in order to show the result more precisely.)



The straight-line path followed by a helicopter between the two points is shorter than the 14 blocks walked by the pedestrian. All blocks are square and the same size.

The fact that the straight-line distance (10.3 blocks) in [\[link\]](#) is less than the total distance walked (14 blocks) is one example of a general characteristic of vectors. (Recall that **vectors** are quantities that have both magnitude and direction.)

As for one-dimensional kinematics, we use arrows to represent vectors. The length of the arrow is proportional to the vector's magnitude. The arrow's length is indicated by hash marks in [\[link\]](#) and [\[link\]](#). The arrow points in the same direction as the vector. For two-dimensional motion, the path of an object can be represented with three vectors: one vector shows the straight-line path between the initial and final points of the motion, one vector shows the horizontal component of the motion, and one vector shows the vertical component of the motion. The horizontal and vertical components of the motion add together to give the straight-line path. For example, observe the three vectors in [\[link\]](#). The first represents a 9-block displacement east. The second represents a 5-block displacement north. These vectors are added to give the third vector, with a 10.3-block total displacement. The third vector is the straight-line path between the two points. Note that in this example, the vectors that we are adding are perpendicular to each other and thus form a right triangle. This means that we can use the Pythagorean theorem to calculate the magnitude of the total displacement. (Note that we cannot use the Pythagorean theorem to add vectors that are not perpendicular. We will develop techniques for adding vectors having any direction, not just those perpendicular to one another, in [Vector Addition and Subtraction: Graphical Methods](#) and [Vector Addition and Subtraction: Analytical Methods](#).)

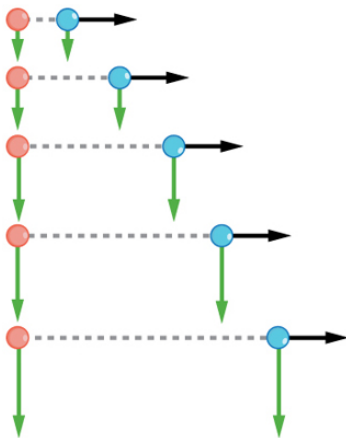
The Independence of Perpendicular Motions

The person taking the path shown in [\[link\]](#) walks east and then north (two perpendicular directions). How far he or she walks east is only affected by his or her motion eastward. Similarly, how far he or she walks north is only affected by his or her motion northward.

Note:**Independence of Motion**

The horizontal and vertical components of two-dimensional motion are independent of each other. Any motion in the horizontal direction does not affect motion in the vertical direction, and vice versa.

This is true in a simple scenario like that of walking in one direction first, followed by another. It is also true of more complicated motion involving movement in two directions at once. For example, let's compare the motions of two baseballs. One baseball is dropped from rest. At the same instant, another is thrown horizontally from the same height and follows a curved path. A stroboscope has captured the positions of the balls at fixed time intervals as they fall.



This shows the motions of two identical balls—one falls from rest, the other has an initial horizontal velocity. Each subsequent

position is an equal time interval. Arrows represent horizontal and vertical velocities at each position. The ball on the right has an initial horizontal velocity, while the ball on the left has no horizontal velocity. Despite the difference in horizontal velocities, the vertical velocities and positions are identical for both balls. This shows that the vertical and horizontal motions are independent.

It is remarkable that for each flash of the strobe, the vertical positions of the two balls are the same. This similarity implies that the vertical motion is independent of whether or not the ball is moving horizontally. (Assuming no air resistance, the vertical motion of a falling object is influenced by gravity only, and not by any horizontal forces.) Careful examination of the

ball thrown horizontally shows that it travels the same horizontal distance between flashes. This is due to the fact that there are no additional forces on the ball in the horizontal direction after it is thrown. This result means that the horizontal velocity is constant, and affected neither by vertical motion nor by gravity (which is vertical). Note that this case is true only for ideal conditions. In the real world, air resistance will affect the speed of the balls in both directions.

The two-dimensional curved path of the horizontally thrown ball is composed of two independent one-dimensional motions (horizontal and vertical). The key to analyzing such motion, called *projectile motion*, is to *resolve* (break) it into motions along perpendicular directions. Resolving two-dimensional motion into perpendicular components is possible because the components are independent. We shall see how to resolve vectors in [Vector Addition and Subtraction: Graphical Methods](#) and [Vector Addition and Subtraction: Analytical Methods](#). We will find such techniques to be useful in many areas of physics.

Note:**PhET Explorations: Ladybug Motion 2D**

Learn about position, velocity and acceleration vectors. Move the ladybug by setting the position, velocity or acceleration, and see how the vectors change. Choose linear, circular or elliptical motion, and record and playback the motion to analyze the behavior.

<https://archive.cnx.org/specials/317a2b1e-2fbd-11e5-99b5-e38ffb545fe6/ladybug-motion/#sim-ladybug-motion>

Summary

- The shortest path between any two points is a straight line. In two dimensions, this path can be represented by a vector with horizontal and vertical components.
- The horizontal and vertical components of a vector are independent of one another. Motion in the horizontal direction does not affect motion

in the vertical direction, and vice versa.

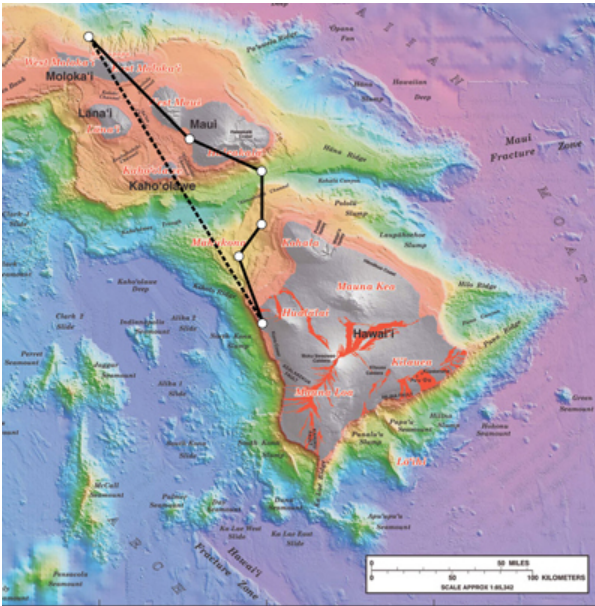
Glossary

vector

a quantity that has both magnitude and direction; an arrow used to represent quantities with both magnitude and direction

Vector Addition and Subtraction: Graphical Methods

- Understand the rules of vector addition, subtraction, and multiplication.
- Apply graphical methods of vector addition and subtraction to determine the displacement of moving objects.



Displacement can be determined graphically using a scale map, such as this one of the Hawaiian Islands. A journey from Hawai'i to Moloka'i has a number of legs, or journey segments. These segments can be added graphically with a ruler to determine the total two-dimensional displacement of the journey. (credit: US Geological Survey)

Vectors in Two Dimensions

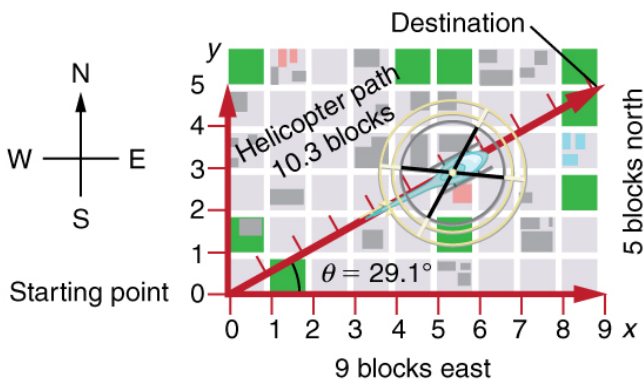
A **vector** is a quantity that has magnitude and direction. Displacement, velocity, acceleration, and force, for example, are all vectors. In one-dimensional, or straight-line, motion, the direction of a vector can be given simply by a plus or minus sign. In two dimensions (2-d), however, we specify the direction of a vector relative to some reference frame (i.e., coordinate system), using an arrow having length proportional to the vector's magnitude and pointing in the direction of the vector.

[\[link\]](#) shows such a *graphical representation of a vector*, using as an example the total displacement for the person walking in a city considered in [Kinematics in Two Dimensions: An Introduction](#). We shall use the notation that a boldface symbol, such as \mathbf{D} , stands for a vector. Its magnitude is represented by the symbol in italics, D , and its direction by θ .

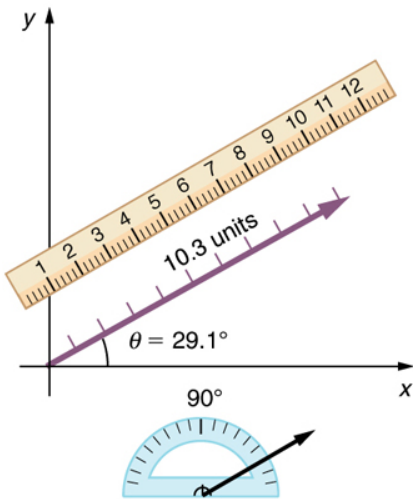
Note:

Vectors in this Text

In this text, we will represent a vector with a boldface variable. For example, we will represent the quantity force with the vector \mathbf{F} , which has both magnitude and direction. The magnitude of the vector will be represented by a variable in italics, such as F , and the direction of the variable will be given by an angle θ .



A person walks 9 blocks east and 5 blocks north. The displacement is 10.3 blocks at an angle 29.1° north of east.



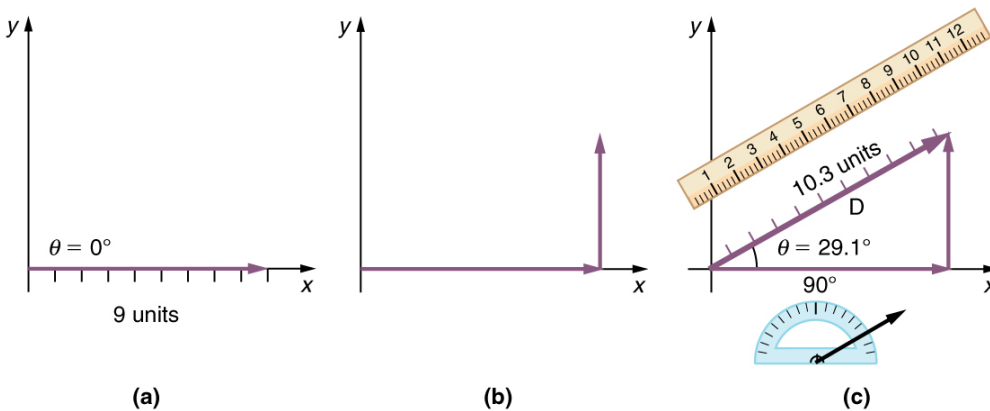
To describe the resultant vector for the person walking in a city considered in [\[link\]](#) graphically, draw an arrow to represent the total displacement vector

D. Using a protractor, draw a line at an angle θ relative to the east-west axis. The length D of the arrow is proportional to the vector's

magnitude and is measured along the line with a ruler. In this example, the magnitude D of the vector is 10.3 units, and the direction θ is 29.1° north of east.

Vector Addition: Head-to-Tail Method

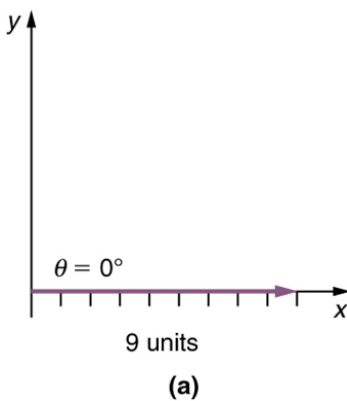
The **head-to-tail method** is a graphical way to add vectors, described in [\[link\]](#) below and in the steps following. The **tail** of the vector is the starting point of the vector, and the **head** (or tip) of a vector is the final, pointed end of the arrow.



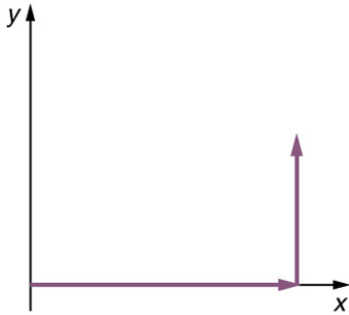
Head-to-Tail Method: The head-to-tail method of graphically adding vectors is illustrated for the two displacements of the person walking in a city considered in [\[link\]](#). (a) Draw a vector representing the displacement to the east. (b) Draw a vector representing the displacement to the north. The tail of this vector should originate from the head of the first, east-pointing vector.

(c) Draw a line from the tail of the east-pointing vector to the head of the north-pointing vector to form the sum or **resultant vector D**. The length of the arrow D is proportional to the vector's magnitude and is measured to be 10.3 units . Its direction, described as the angle with respect to the east (or horizontal axis) θ is measured with a protractor to be 29.1° .

Step 1. Draw an arrow to represent the first vector (9 blocks to the east) using a ruler and protractor.



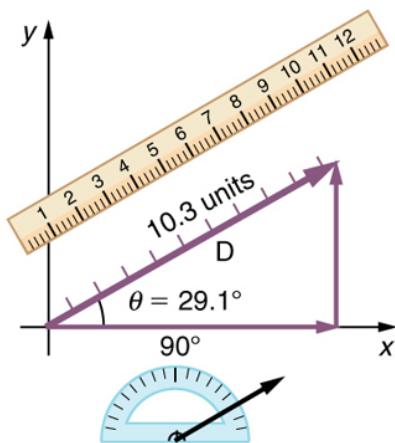
Step 2. Now draw an arrow to represent the second vector (5 blocks to the north). Place the tail of the second vector at the head of the first vector.



(b)

Step 3. If there are more than two vectors, continue this process for each vector to be added. Note that in our example, we have only two vectors, so we have finished placing arrows tip to tail.

Step 4. Draw an arrow from the tail of the first vector to the head of the last vector. This is the **resultant**, or the sum, of the other vectors.



(c)

Step 5. To get the **magnitude** of the resultant, measure its length with a ruler. (Note that in most calculations, we will use the Pythagorean theorem to determine this length.)

Step 6. To get the **direction** of the resultant, measure the angle it makes with the reference frame using a protractor. (Note that in most calculations, we will use trigonometric relationships to determine this angle.)

The graphical addition of vectors is limited in accuracy only by the precision with which the drawings can be made and the precision of the measuring tools. It is valid for any number of vectors.

Example:

Adding Vectors Graphically Using the Head-to-Tail Method: A Woman Takes a Walk

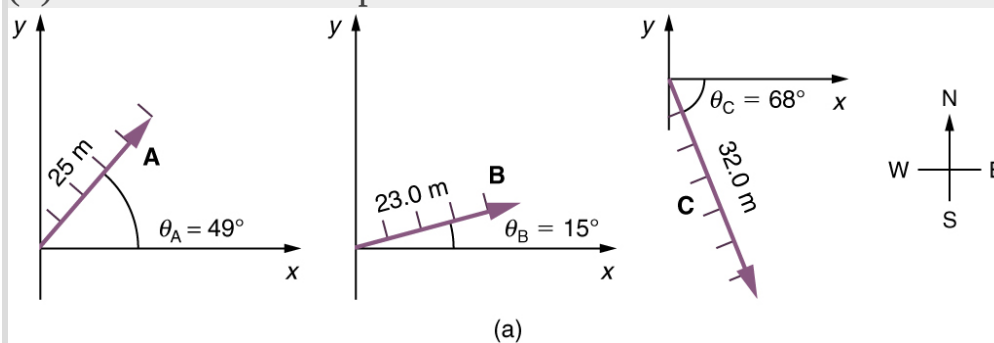
Use the graphical technique for adding vectors to find the total displacement of a person who walks the following three paths (displacements) on a flat field. First, she walks 25.0 m in a direction 49.0° north of east. Then, she walks 23.0 m heading 15.0° north of east. Finally, she turns and walks 32.0 m in a direction 68.0° south of east.

Strategy

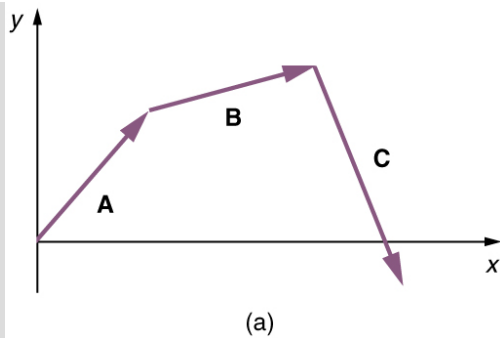
Represent each displacement vector graphically with an arrow, labeling the first **A**, the second **B**, and the third **C**, making the lengths proportional to the distance and the directions as specified relative to an east-west line. The head-to-tail method outlined above will give a way to determine the magnitude and direction of the resultant displacement, denoted **R**.

Solution

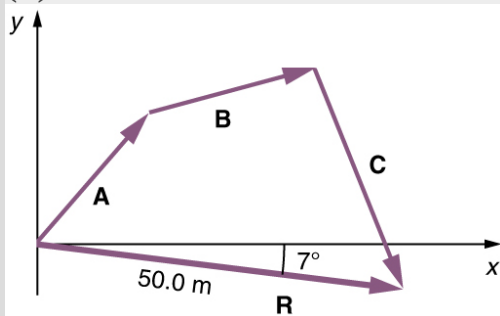
(1) Draw the three displacement vectors.



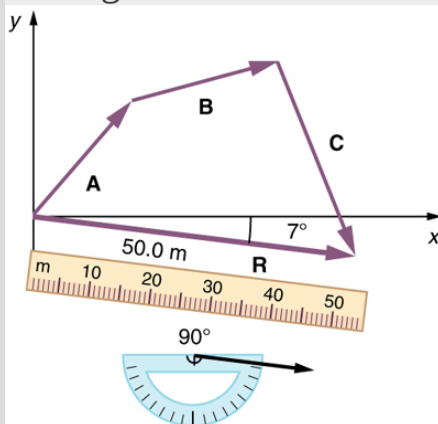
(2) Place the vectors head to tail retaining both their initial magnitude and direction.



(3) Draw the resultant vector, \mathbf{R} .



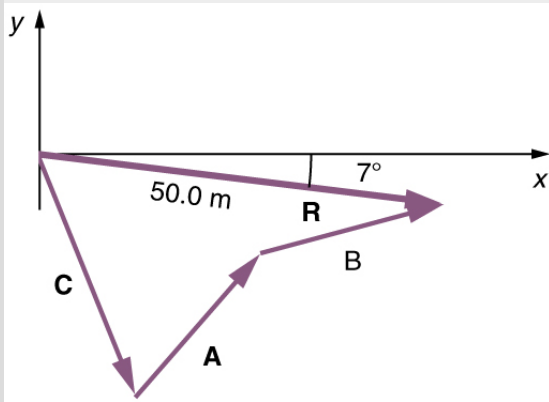
(4) Use a ruler to measure the magnitude of \mathbf{R} , and a protractor to measure the direction of \mathbf{R} . While the direction of the vector can be specified in many ways, the easiest way is to measure the angle between the vector and the nearest horizontal or vertical axis. Since the resultant vector is south of the eastward pointing axis, we flip the protractor upside down and measure the angle between the eastward axis and the vector.



In this case, the total displacement \mathbf{R} is seen to have a magnitude of 50.0 m and to lie in a direction 7.0° south of east. By using its magnitude and direction, this vector can be expressed as $R = 50.0 \text{ m}$ and $\theta = 7.0^\circ$ south of east.

Discussion

The head-to-tail graphical method of vector addition works for any number of vectors. It is also important to note that the resultant is independent of the order in which the vectors are added. Therefore, we could add the vectors in any order as illustrated in [\[link\]](#) and we will still get the same solution.



Here, we see that when the same vectors are added in a different order, the result is the same. This characteristic is true in every case and is an important characteristic of vectors. Vector addition is **commutative**. Vectors can be added in any order.

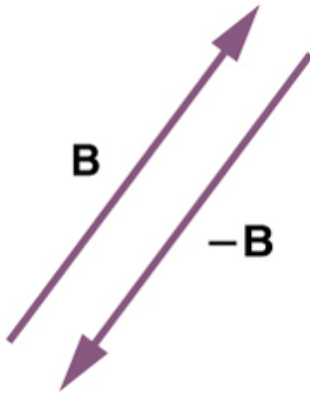
Equation:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}.$$

(This is true for the addition of ordinary numbers as well—you get the same result whether you add $2 + 3$ or $3 + 2$, for example).

Vector Subtraction

Vector subtraction is a straightforward extension of vector addition. To define subtraction (say we want to subtract \mathbf{B} from \mathbf{A} , written $\mathbf{A} - \mathbf{B}$, we must first define what we mean by subtraction. The *negative* of a vector \mathbf{B} is defined to be $-\mathbf{B}$; that is, graphically *the negative of any vector has the same magnitude but the opposite direction*, as shown in [\[link\]](#). In other words, \mathbf{B} has the same length as $-\mathbf{B}$, but points in the opposite direction. Essentially, we just flip the vector so it points in the opposite direction.



The negative of a vector is just another vector of the same magnitude but pointing in the opposite direction. So **B** is the negative of **-B**; it has the same length but opposite direction.

The *subtraction* of vector **B** from vector **A** is then simply defined to be the addition of **-B** to **A**. Note that vector subtraction is the addition of a negative vector. The order of subtraction does not affect the results.

Equation:

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}).$$

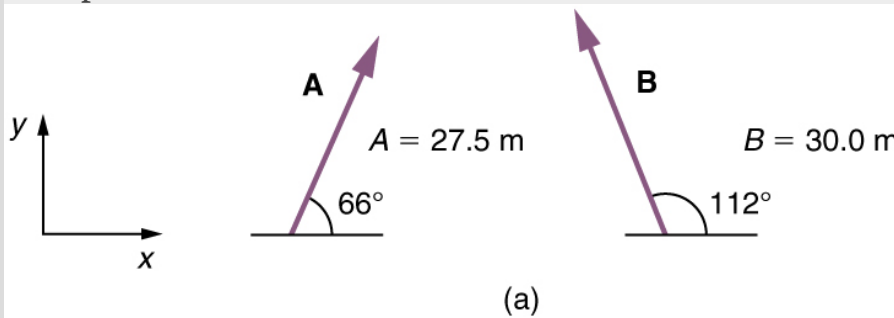
This is analogous to the subtraction of scalars (where, for example, $5 - 2 = 5 + (-2)$). Again, the result is independent of the order in which

the subtraction is made. When vectors are subtracted graphically, the techniques outlined above are used, as the following example illustrates.

Example:

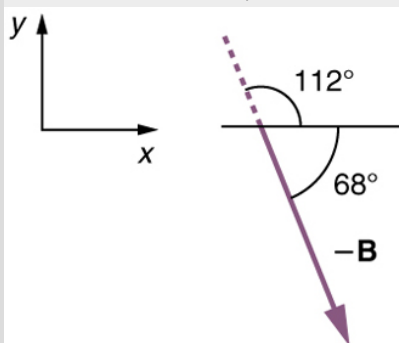
Subtracting Vectors Graphically: A Woman Sailing a Boat

A woman sailing a boat at night is following directions to a dock. The instructions read to first sail 27.5 m in a direction 66.0° north of east from her current location, and then travel 30.0 m in a direction 112° north of east (or 22.0° west of north). If the woman makes a mistake and travels in the *opposite* direction for the second leg of the trip, where will she end up? Compare this location with the location of the dock.



Strategy

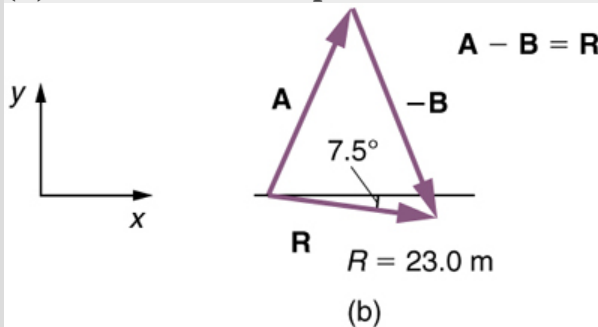
We can represent the first leg of the trip with a vector **A**, and the second leg of the trip with a vector **B**. The dock is located at a location $\mathbf{A} + \mathbf{B}$. If the woman mistakenly travels in the *opposite* direction for the second leg of the journey, she will travel a distance B (30.0 m) in the direction $180^\circ - 112^\circ = 68^\circ$ south of east. We represent this as $-\mathbf{B}$, as shown below. The vector $-\mathbf{B}$ has the same magnitude as **B** but is in the opposite direction. Thus, she will end up at a location $\mathbf{A} + (-\mathbf{B})$, or $\mathbf{A} - \mathbf{B}$.



We will perform vector addition to compare the location of the dock, $\mathbf{A} + \mathbf{B}$, with the location at which the woman mistakenly arrives, $\mathbf{A} + (-\mathbf{B})$.

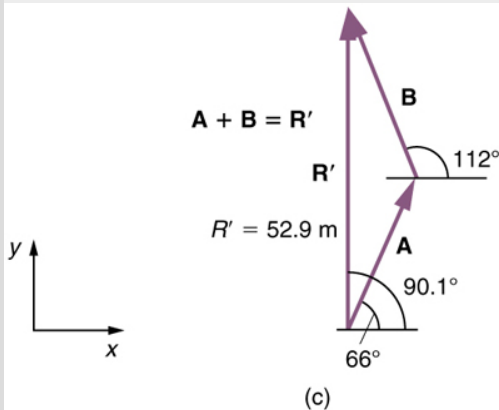
Solution

- (1) To determine the location at which the woman arrives by accident, draw vectors \mathbf{A} and $-\mathbf{B}$.
- (2) Place the vectors head to tail.
- (3) Draw the resultant vector \mathbf{R} .
- (4) Use a ruler and protractor to measure the magnitude and direction of \mathbf{R} .



In this case, $R = 23.0 \text{ m}$ and $\theta = 7.5^\circ$ south of east.

- (5) To determine the location of the dock, we repeat this method to add vectors \mathbf{A} and \mathbf{B} . We obtain the resultant vector \mathbf{R}' :



In this case $R = 52.9 \text{ m}$ and $\theta = 90.1^\circ$ north of east.

We can see that the woman will end up a significant distance from the dock if she travels in the opposite direction for the second leg of the trip.

Discussion

Because subtraction of a vector is the same as addition of a vector with the opposite direction, the graphical method of subtracting vectors works the same as for addition.

Multiplication of Vectors and Scalars

If we decided to walk three times as far on the first leg of the trip considered in the preceding example, then we would walk 3×27.5 m, or 82.5 m, in a direction 66.0° north of east. This is an example of multiplying a vector by a positive **scalar**. Notice that the magnitude changes, but the direction stays the same.

If the scalar is negative, then multiplying a vector by it changes the vector's magnitude and gives the new vector the *opposite* direction. For example, if you multiply by -2 , the magnitude doubles but the direction changes. We can summarize these rules in the following way: When vector **A** is multiplied by a scalar c ,

- the magnitude of the vector becomes the absolute value of cA ,
- if c is positive, the direction of the vector does not change,
- if c is negative, the direction is reversed.

In our case, $c = 3$ and $A = 27.5$ m. Vectors are multiplied by scalars in many situations. Note that division is the inverse of multiplication. For example, dividing by 2 is the same as multiplying by the value $(1/2)$. The rules for multiplication of vectors by scalars are the same for division; simply treat the divisor as a scalar between 0 and 1.

Resolving a Vector into Components

In the examples above, we have been adding vectors to determine the resultant vector. In many cases, however, we will need to do the opposite. We will need to take a single vector and find what other vectors added together produce it. In most cases, this involves determining the perpendicular **components** of a single vector, for example the x - and y -components, or the north-south and east-west components.

For example, we may know that the total displacement of a person walking in a city is 10.3 blocks in a direction 29.0° north of east and want to find out how many blocks east and north had to be walked. This method is called *finding the components (or parts)* of the displacement in the east and north

directions, and it is the inverse of the process followed to find the total displacement. It is one example of finding the components of a vector. There are many applications in physics where this is a useful thing to do. We will see this soon in [Projectile Motion](#), and much more when we cover **forces** in [Dynamics: Newton's Laws of Motion](#). Most of these involve finding components along perpendicular axes (such as north and east), so that right triangles are involved. The analytical techniques presented in [Vector Addition and Subtraction: Analytical Methods](#) are ideal for finding vector components.

Note:

PhET Explorations: Maze Game

Learn about position, velocity, and acceleration in the "Arena of Pain". Use the green arrow to move the ball. Add more walls to the arena to make the game more difficult. Try to make a goal as fast as you can.

<https://archive.cnx.org/specials/30e37034-2fbd-11e5-83a2-03be60006ece/maze-game/#sim-maze-game>

Summary

- The **graphical method of adding vectors \mathbf{A} and \mathbf{B}** involves drawing vectors on a graph and adding them using the head-to-tail method. The resultant vector \mathbf{R} is defined such that $\mathbf{A} + \mathbf{B} = \mathbf{R}$. The magnitude and direction of \mathbf{R} are then determined with a ruler and protractor, respectively.
- The **graphical method of subtracting vector \mathbf{B} from \mathbf{A}** involves adding the opposite of vector \mathbf{B} , which is defined as $-\mathbf{B}$. In this case, $\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}) = \mathbf{R}$. Then, the head-to-tail method of addition is followed in the usual way to obtain the resultant vector \mathbf{R} .
- Addition of vectors is **commutative** such that $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.
- The **head-to-tail method** of adding vectors involves drawing the first vector on a graph and then placing the tail of each subsequent vector at

the head of the previous vector. The resultant vector is then drawn from the tail of the first vector to the head of the final vector.

- If a vector \mathbf{A} is multiplied by a scalar quantity c , the magnitude of the product is given by cA . If c is positive, the direction of the product points in the same direction as \mathbf{A} ; if c is negative, the direction of the product points in the opposite direction as \mathbf{A} .

Conceptual Questions

Exercise:

Problem:

Which of the following is a vector: a person's height, the altitude on Mt. Everest, the age of the Earth, the boiling point of water, the cost of this book, the Earth's population, the acceleration of gravity?

Exercise:

Problem:

Give a specific example of a vector, stating its magnitude, units, and direction.

Exercise:

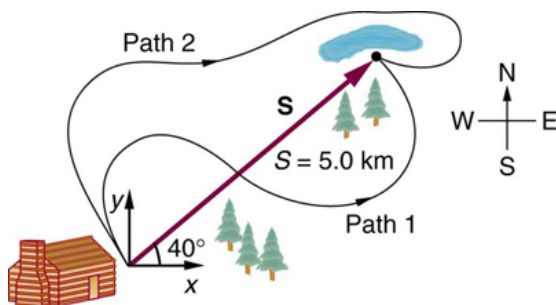
Problem:

What do vectors and scalars have in common? How do they differ?

Exercise:

Problem:

Two campers in a national park hike from their cabin to the same spot on a lake, each taking a different path, as illustrated below. The total distance traveled along Path 1 is 7.5 km, and that along Path 2 is 8.2 km. What is the final displacement of each camper?



Exercise:

Problem:

If an airplane pilot is told to fly 123 km in a straight line to get from San Francisco to Sacramento, explain why he could end up anywhere on the circle shown in [\[link\]](#). What other information would he need to get to Sacramento?



Exercise:

Problem:

Suppose you take two steps **A** and **B** (that is, two nonzero displacements). Under what circumstances can you end up at your starting point? More generally, under what circumstances can two nonzero vectors add to give zero? Is the maximum distance you can end up from the starting point $\mathbf{A} + \mathbf{B}$ the sum of the lengths of the two steps?

Exercise:

Problem: Explain why it is not possible to add a scalar to a vector.

Exercise:

Problem:

If you take two steps of different sizes, can you end up at your starting point? More generally, can two vectors with different magnitudes ever add to zero? Can three or more?

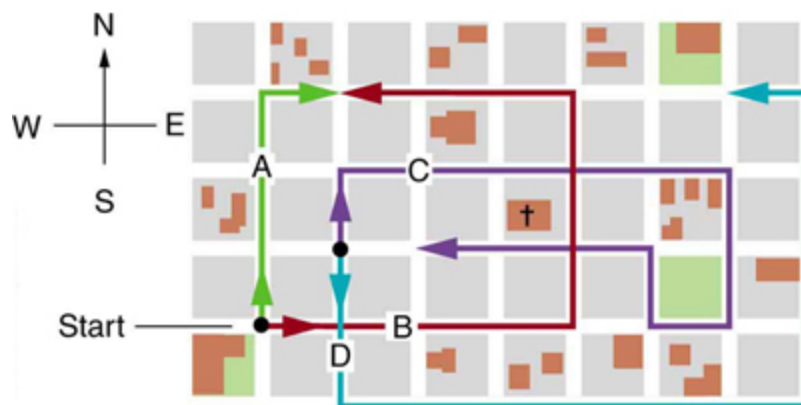
Problems & Exercises

Use graphical methods to solve these problems. You may assume data taken from graphs is accurate to three digits.

Exercise:

Problem:

Find the following for path A in [\[link\]](#): (a) the total distance traveled, and (b) the magnitude and direction of the displacement from start to finish.



The various lines represent paths taken by different people walking in a city. All blocks are 120 m on a side.

Solution:

(a) 480 m

(b) 379 m, 18.4° east of north

Exercise:**Problem:**

Find the following for path B in [\[link\]](#): (a) the total distance traveled, and (b) the magnitude and direction of the displacement from start to finish.

Exercise:**Problem:**

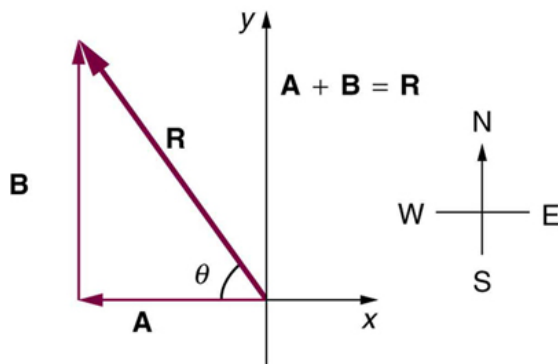
Find the north and east components of the displacement for the hikers shown in [\[link\]](#).

Solution:

north component 3.21 km, east component 3.83 km

Exercise:**Problem:**

Suppose you walk 18.0 m straight west and then 25.0 m straight north. How far are you from your starting point, and what is the compass direction of a line connecting your starting point to your final position? (If you represent the two legs of the walk as vector displacements **A** and **B**, as in [\[link\]](#), then this problem asks you to find their sum $\mathbf{R} = \mathbf{A} + \mathbf{B}$.)

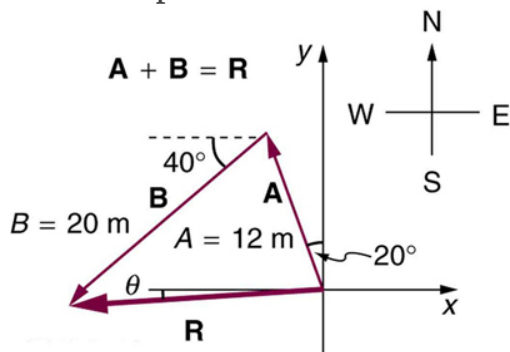


The two displacements **A** and **B** add to give a total displacement **R** having magnitude R and direction θ .

Exercise:

Problem:

Suppose you first walk 12.0 m in a direction 20° west of north and then 20.0 m in a direction 40.0° south of west. How far are you from your starting point, and what is the compass direction of a line connecting your starting point to your final position? (If you represent the two legs of the walk as vector displacements **A** and **B**, as in [\[link\]](#), then this problem finds their sum $\mathbf{R} = \mathbf{A} + \mathbf{B}$.)



Solution:

19.5 m, 4.65° south of west

Exercise:**Problem:**

Repeat the problem above, but reverse the order of the two legs of the walk; show that you get the same final result. That is, you first walk leg **B**, which is 20.0 m in a direction exactly 40° south of west, and then leg **A**, which is 12.0 m in a direction exactly 20° west of north. (This problem shows that $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.)

Exercise:**Problem:**

(a) Repeat the problem two problems prior, but for the second leg you walk 20.0 m in a direction 40.0° north of east (which is equivalent to subtracting **B** from **A**—that is, to finding $\mathbf{R} = \mathbf{A} - \mathbf{B}$). (b) Repeat the problem two problems prior, but now you first walk 20.0 m in a direction 40.0° south of west and then 12.0 m in a direction 20.0° east of south (which is equivalent to subtracting **A** from **B**—that is, to finding $\mathbf{R} = \mathbf{B} - \mathbf{A} = -\mathbf{R}'$). Show that this is the case.

Solution:

(a) 26.6 m, 65.1° north of east

(b) 26.6 m, 65.1° south of west

Exercise:**Problem:**

Show that the *order* of addition of three vectors does not affect their sum. Show this property by choosing any three vectors **A**, **B**, and **C**, all having different lengths and directions. Find the sum $\mathbf{A} + \mathbf{B} + \mathbf{C}$ then find their sum when added in a different order and show the result is the same. (There are five other orders in which **A**, **B**, and **C** can be added; choose only one.)

Exercise:

Problem:

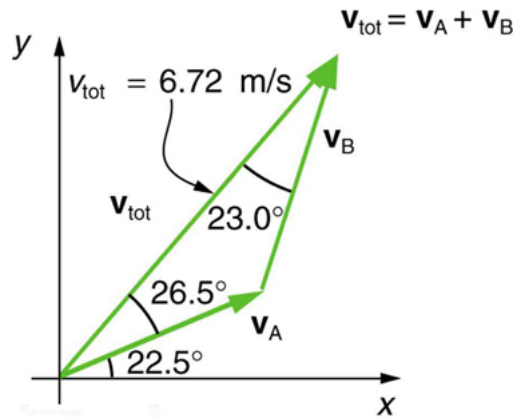
Show that the sum of the vectors discussed in [\[link\]](#) gives the result shown in [\[link\]](#).

Solution:

52.9 m, 90.1° with respect to the x -axis.

Exercise:

Problem: Find the magnitudes of velocities v_A and v_B in [\[link\]](#)



The two velocities \mathbf{v}_A
and \mathbf{v}_B add to give a total
 \mathbf{v}_{tot} .

Exercise:**Problem:**

Find the components of v_{tot} along the x - and y -axes in [\[link\]](#).

Solution:

x -component 4.41 m/s

y-component 5.07 m/s

Exercise:

Problem:

Find the components of v_{tot} along a set of perpendicular axes rotated 30° counterclockwise relative to those in [\[link\]](#).

Glossary

component (of a 2-d vector)

a piece of a vector that points in either the vertical or the horizontal direction; every 2-d vector can be expressed as a sum of two vertical and horizontal vector components

commutative

refers to the interchangeability of order in a function; vector addition is commutative because the order in which vectors are added together does not affect the final sum

direction (of a vector)

the orientation of a vector in space

head (of a vector)

the end point of a vector; the location of the tip of the vector's arrowhead; also referred to as the "tip"

head-to-tail method

a method of adding vectors in which the tail of each vector is placed at the head of the previous vector

magnitude (of a vector)

the length or size of a vector; magnitude is a scalar quantity

resultant

the sum of two or more vectors

resultant vector

the vector sum of two or more vectors

scalar

a quantity with magnitude but no direction

tail

the start point of a vector; opposite to the head or tip of the arrow

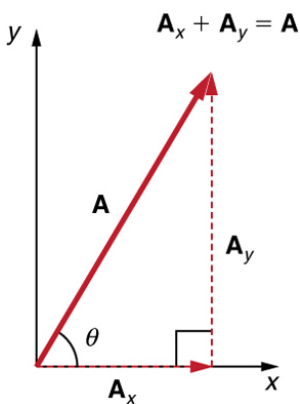
Vector Addition and Subtraction: Analytical Methods

- Understand the rules of vector addition and subtraction using analytical methods.
- Apply analytical methods to determine vertical and horizontal component vectors.
- Apply analytical methods to determine the magnitude and direction of a resultant vector.

Analytical methods of vector addition and subtraction employ geometry and simple trigonometry rather than the ruler and protractor of graphical methods. Part of the graphical technique is retained, because vectors are still represented by arrows for easy visualization. However, analytical methods are more concise, accurate, and precise than graphical methods, which are limited by the accuracy with which a drawing can be made. Analytical methods are limited only by the accuracy and precision with which physical quantities are known.

Resolving a Vector into Perpendicular Components

Analytical techniques and right triangles go hand-in-hand in physics because (among other things) motions along perpendicular directions are independent. We very often need to separate a vector into perpendicular components. For example, given a vector like **A** in [\[link\]](#), we may wish to find which two perpendicular vectors, **A_x** and **A_y**, add to produce it.



The vector \mathbf{A} , with its tail at the origin of an x, y -coordinate system, is shown together with its x - and y -components, \mathbf{A}_x and \mathbf{A}_y . These vectors form a right triangle. The analytical relationships among these vectors are summarized below.

\mathbf{A}_x and \mathbf{A}_y are defined to be the components of \mathbf{A} along the x - and y -axes. The three vectors \mathbf{A} , \mathbf{A}_x , and \mathbf{A}_y form a right triangle:

Equation:

$$\mathbf{A}_x + \mathbf{A}_y = \mathbf{A}.$$

Note that this relationship between vector components and the resultant vector holds only for vector quantities (which include both magnitude and direction). The relationship does not apply for the magnitudes alone. For example, if $\mathbf{A}_x = 3$ m east, $\mathbf{A}_y = 4$ m north, and $\mathbf{A} = 5$ m north-east, then it is true that the vectors $\mathbf{A}_x + \mathbf{A}_y = \mathbf{A}$. However, it is *not* true that the sum of the magnitudes of the vectors is also equal. That is,

Equation:

$$3 \text{ m} + 4 \text{ m} \neq 5 \text{ m}$$

Thus,

Equation:

$$A_x + A_y \neq A$$

If the vector **A** is known, then its magnitude A (its length) and its angle θ (its direction) are known. To find A_x and A_y , its x - and y -components, we use the following relationships for a right triangle.

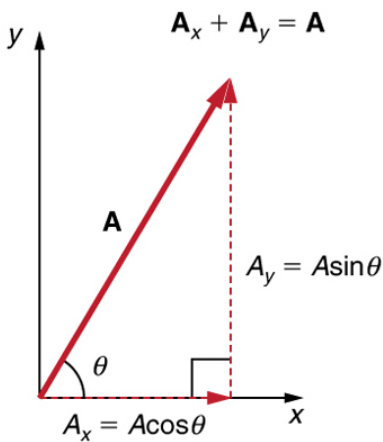
Equation:

$$A_x = A \cos \theta$$

and

Equation:

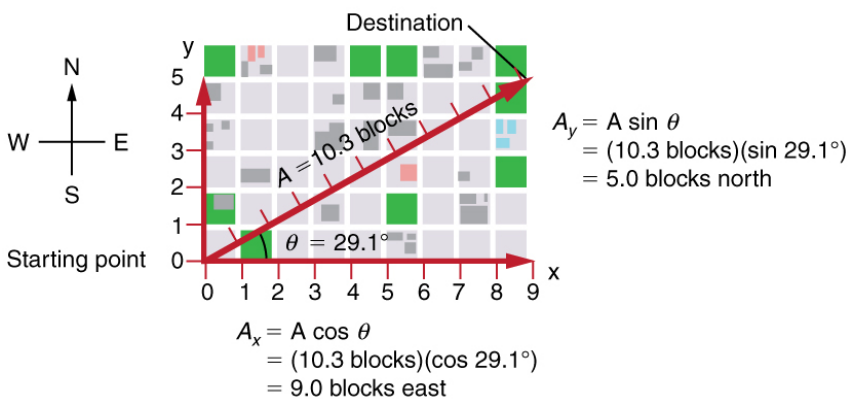
$$A_y = A \sin \theta.$$



The magnitudes of
the vector

components A_x
and A_y can be
related to the
resultant vector A
and the angle θ
with trigonometric
identities. Here we
see that
 $A_x = A \cos \theta$ and
 $A_y = A \sin \theta$.

Suppose, for example, that A is the vector representing the total displacement of the person walking in a city considered in [Kinematics in Two Dimensions: An Introduction](#) and [Vector Addition and Subtraction: Graphical Methods](#).



We can use the relationships $A_x = A \cos \theta$ and
 $A_y = A \sin \theta$ to determine the magnitude of
the horizontal and vertical component vectors
in this example.

Then $A = 10.3$ blocks and $\theta = 29.1^\circ$, so that

Equation:

$$A_x = A \cos \theta = (10.3 \text{ blocks})(\cos 29.1^\circ) = 9.0 \text{ blocks}$$

Equation:

$$A_y = A \sin \theta = (10.3 \text{ blocks})(\sin 29.1^\circ) = 5.0 \text{ blocks.}$$

Calculating a Resultant Vector

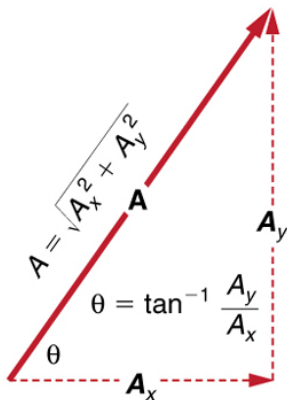
If the perpendicular components \mathbf{A}_x and \mathbf{A}_y of a vector \mathbf{A} are known, then \mathbf{A} can also be found analytically. To find the magnitude A and direction θ of a vector from its perpendicular components \mathbf{A}_x and \mathbf{A}_y , we use the following relationships:

Equation:

$$A = \sqrt{A_x^2 + A_y^2}$$

Equation:

$$\theta = \tan^{-1}(A_y/A_x).$$



The
magnitude
and direction
of the
resultant
vector can
be
determined
once the
horizontal
and vertical
components
 A_x and A_y
have been
determined.

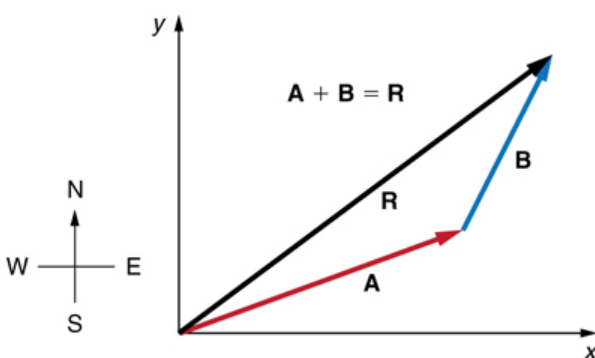
Note that the equation $A = \sqrt{A_x^2 + A_y^2}$ is just the Pythagorean theorem relating the legs of a right triangle to the length of the hypotenuse. For example, if A_x and A_y are 9 and 5 blocks, respectively, then $A = \sqrt{9^2 + 5^2} = 10.3$ blocks, again consistent with the example of the person walking in a city. Finally, the direction is $\theta = \tan^{-1}(5/9) = 29.1^\circ$, as before.

Note:

Determining Vectors and Vector Components with Analytical Methods
Equations $A_x = A \cos \theta$ and $A_y = A \sin \theta$ are used to find the perpendicular components of a vector—that is, to go from A and θ to A_x and A_y . Equations $A = \sqrt{A_x^2 + A_y^2}$ and $\theta = \tan^{-1}(A_y/A_x)$ are used to find a vector from its perpendicular components—that is, to go from A_x and A_y to A and θ . Both processes are crucial to analytical methods of vector addition and subtraction.

Adding Vectors Using Analytical Methods

To see how to add vectors using perpendicular components, consider [\[link\]](#), in which the vectors **A** and **B** are added to produce the resultant **R**.



Vectors **A** and **B** are two legs of a walk, and **R** is the resultant or total displacement. You can use analytical methods to determine the magnitude and direction of **R**.

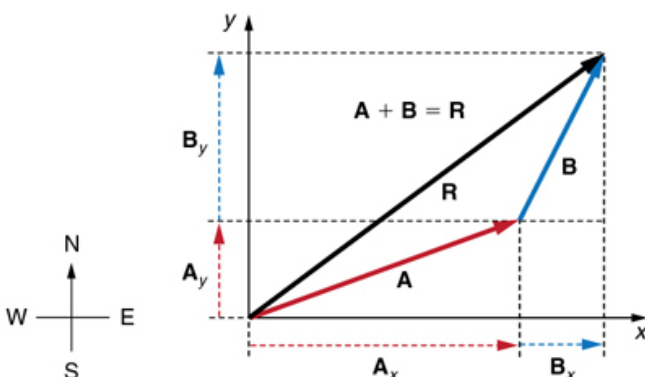
If **A** and **B** represent two legs of a walk (two displacements), then **R** is the total displacement. The person taking the walk ends up at the tip of **R**.

There are many ways to arrive at the same point. In particular, the person could have walked first in the x -direction and then in the y -direction. Those paths are the x - and y -components of the resultant, \mathbf{R}_x and \mathbf{R}_y . If we know

\mathbf{R}_x and \mathbf{R}_y , we can find R and θ using the equations $R = \sqrt{R_x^2 + R_y^2}$ and $\theta = \tan^{-1}(R_y/R_x)$. When you use the analytical method of vector addition, you can determine the components or the magnitude and direction of a vector.

Step 1. Identify the x - and y -axes that will be used in the problem. Then, find the components of each vector to be added along the chosen

perpendicular axes. Use the equations $A_x = A \cos \theta$ and $A_y = A \sin \theta$ to find the components. In [\[link\]](#), these components are A_x , A_y , B_x , and B_y . The angles that vectors **A** and **B** make with the x -axis are θ_A and θ_B , respectively.



To add vectors **A** and **B**, first determine the horizontal and vertical components of each vector. These are the dotted vectors A_x , A_y , B_x and B_y shown in the image.

Step 2. Find the components of the resultant along each axis by adding the components of the individual vectors along that axis. That is, as shown in [\[link\]](#),

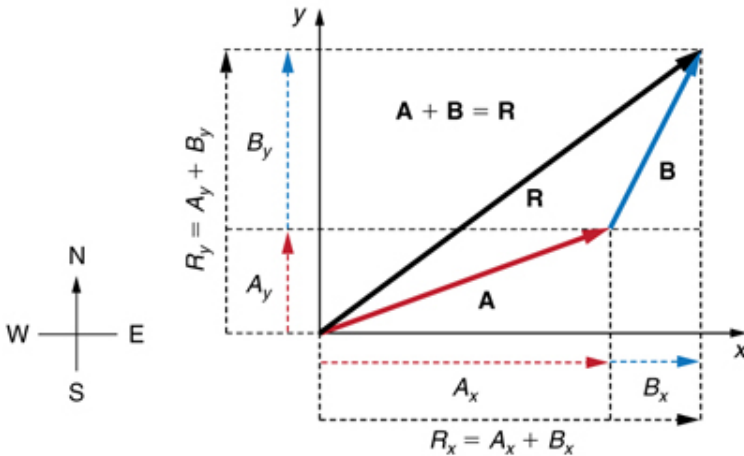
Equation:

$$R_x = A_x + B_x$$

and

Equation:

$$R_y = A_y + B_y.$$



The magnitude of the vectors \mathbf{A}_x and \mathbf{B}_x add to give the magnitude R_x of the resultant vector in the horizontal direction. Similarly, the magnitudes of the vectors \mathbf{A}_y and \mathbf{B}_y add to give the magnitude R_y of the resultant vector in the vertical direction.

Components along the same axis, say the x-axis, are vectors along the same line and, thus, can be added to one another like ordinary numbers. The same is true for components along the y-axis. (For example, a 9-block eastward walk could be taken in two legs, the first 3 blocks east and the second 6 blocks east, for a total of 9, because they are along the same direction.) So resolving vectors into components along common axes makes it easier to add them. Now that the components of \mathbf{R} are known, its magnitude and direction can be found.

Step 3. To get the magnitude R of the resultant, use the Pythagorean theorem:

Equation:

$$R = \sqrt{R_x^2 + R_y^2}.$$

Step 4. To get the direction of the resultant:

Equation:

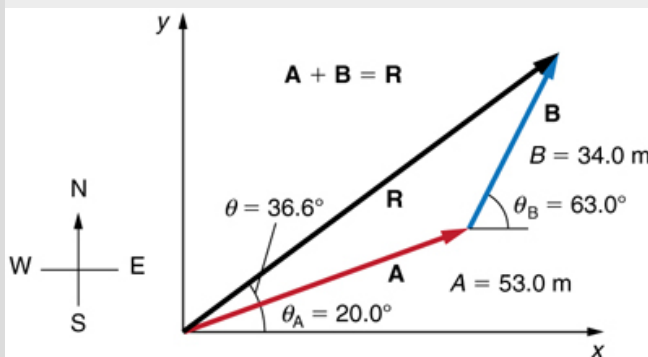
$$\theta = \tan^{-1}(R_y/R_x).$$

The following example illustrates this technique for adding vectors using perpendicular components.

Example:

Adding Vectors Using Analytical Methods

Add the vector **A** to the vector **B** shown in [\[link\]](#), using perpendicular components along the *x*- and *y*-axes. The *x*- and *y*-axes are along the east–west and north–south directions, respectively. Vector **A** represents the first leg of a walk in which a person walks 53.0 m in a direction 20.0° north of east. Vector **B** represents the second leg, a displacement of 34.0 m in a direction 63.0° north of east.



Vector **A** has magnitude 53.0 m and direction 20.0° north of the *x*-axis. Vector **B** has magnitude 34.0 m and direction 63.0° north of the *x*-axis. You can use analytical methods to determine the magnitude and direction of **R**.

Strategy

The components of **A** and **B** along the x - and y -axes represent walking due east and due north to get to the same ending point. Once found, they are combined to produce the resultant.

Solution

Following the method outlined above, we first find the components of **A** and **B** along the x - and y -axes. Note that $A = 53.0$ m, $\theta_A = 20.0^\circ$, $B = 34.0$ m, and $\theta_B = 63.0^\circ$. We find the x -components by using $A_x = A \cos \theta$, which gives

Equation:

$$\begin{aligned} A_x &= A \cos \theta_A = (53.0 \text{ m})(\cos 20.0^\circ) \\ &= (53.0 \text{ m})(0.940) = 49.8 \text{ m} \end{aligned}$$

and

Equation:

$$\begin{aligned} B_x &= B \cos \theta_B = (34.0 \text{ m})(\cos 63.0^\circ) \\ &= (34.0 \text{ m})(0.454) = 15.4 \text{ m}. \end{aligned}$$

Similarly, the y -components are found using $A_y = A \sin \theta_A$:

Equation:

$$\begin{aligned} A_y &= A \sin \theta_A = (53.0 \text{ m})(\sin 20.0^\circ) \\ &= (53.0 \text{ m})(0.342) = 18.1 \text{ m} \end{aligned}$$

and

Equation:

$$\begin{aligned} B_y &= B \sin \theta_B = (34.0 \text{ m})(\sin 63.0^\circ) \\ &= (34.0 \text{ m})(0.891) = 30.3 \text{ m}. \end{aligned}$$

The x - and y -components of the resultant are thus

Equation:

$$R_x = A_x + B_x = 49.8 \text{ m} + 15.4 \text{ m} = 65.2 \text{ m}$$

and

Equation:

$$R_y = A_y + B_y = 18.1 \text{ m} + 30.3 \text{ m} = 48.4 \text{ m}.$$

Now we can find the magnitude of the resultant by using the Pythagorean theorem:

Equation:

$$R = \sqrt{R_x^2 + R_y^2} = \sqrt{(65.2)^2 + (48.4)^2} \text{ m}$$

so that

Equation:

$$R = 81.2 \text{ m}.$$

Finally, we find the direction of the resultant:

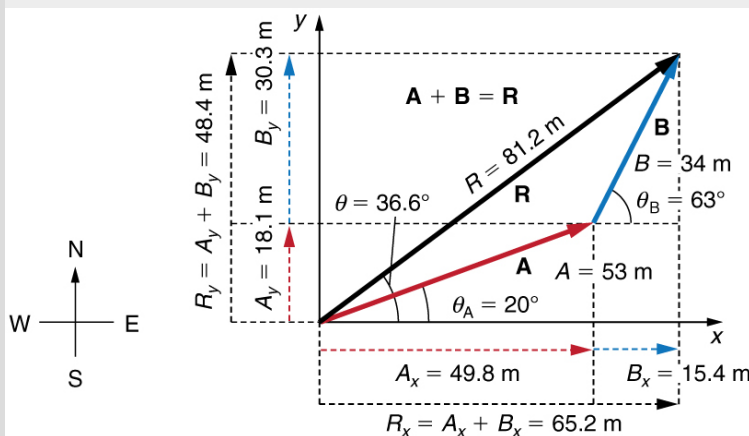
Equation:

$$\theta = \tan^{-1}(R_y/R_x) = +\tan^{-1}(48.4/65.2).$$

Thus,

Equation:

$$\theta = \tan^{-1}(0.742) = 36.6^\circ.$$



Using analytical methods, we see that the magnitude of \mathbf{R} is 81.2 m and its

direction is 36.6° north of east.

Discussion

This example illustrates the addition of vectors using perpendicular components. Vector subtraction using perpendicular components is very similar—it is just the addition of a negative vector.

Subtraction of vectors is accomplished by the addition of a negative vector.

That is, $\mathbf{A} - \mathbf{B} \equiv \mathbf{A} + (-\mathbf{B})$. Thus, *the method for the subtraction of vectors using perpendicular components is identical to that for addition.*

The components of $-\mathbf{B}$ are the negatives of the components of \mathbf{B} . The x- and y-components of the resultant $\mathbf{A} - \mathbf{B} = \mathbf{R}$ are thus

Equation:

$$R_x = A_x + (-B_x)$$

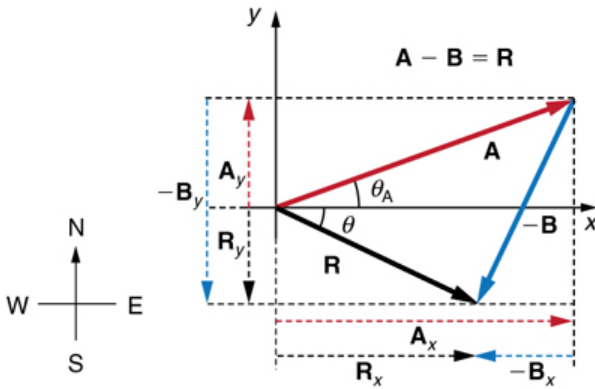
and

Equation:

$$R_y = A_y + (-B_y)$$

and the rest of the method outlined above is identical to that for addition. (See [\[link\]](#).)

Analyzing vectors using perpendicular components is very useful in many areas of physics, because perpendicular quantities are often independent of one another. The next module, [Projectile Motion](#), is one of many in which using perpendicular components helps make the picture clear and simplifies the physics.



The subtraction of the two vectors shown in [\[link\]](#). The components of $-\mathbf{B}$ are the negatives of the components of \mathbf{B} . The method of subtraction is the same as that for addition.

Note:

PhET Explorations: Vector Addition

Learn how to add vectors. Drag vectors onto a graph, change their length and angle, and sum them together. The magnitude, angle, and components of each vector can be displayed in several formats.

https://phet.colorado.edu/sims/vector-addition/vector-addition_en.html

Summary

- The analytical method of vector addition and subtraction involves using the Pythagorean theorem and trigonometric identities to determine the magnitude and direction of a resultant vector.
- The steps to add vectors \mathbf{A} and \mathbf{B} using the analytical method are as follows:

Step 1: Determine the coordinate system for the vectors. Then, determine the horizontal and vertical components of each vector using the equations

Equation:

$$\begin{aligned}A_x &= A \cos \theta \\B_x &= B \cos \theta\end{aligned}$$

and

Equation:

$$\begin{aligned}A_y &= A \sin \theta \\B_y &= B \sin \theta.\end{aligned}$$

Step 2: Add the horizontal and vertical components of each vector to determine the components R_x and R_y of the resultant vector, **R**:

Equation:

$$R_x = A_x + B_x$$

and

Equation:

$$R_y = A_y + B_y.$$

Step 3: Use the Pythagorean theorem to determine the magnitude, R , of the resultant vector **R**:

Equation:

$$R = \sqrt{R_x^2 + R_y^2}.$$

Step 4: Use a trigonometric identity to determine the direction, θ , of **R** :

Equation:

$$\theta = \tan^{-1}(R_y/R_x).$$

Conceptual Questions

Exercise:

Problem:

Suppose you add two vectors **A** and **B**. What relative direction between them produces the resultant with the greatest magnitude? What is the maximum magnitude? What relative direction between them produces the resultant with the smallest magnitude? What is the minimum magnitude?

Exercise:

Problem:

Give an example of a nonzero vector that has a component of zero.

Exercise:

Problem:

Explain why a vector cannot have a component greater than its own magnitude.

Exercise:

Problem:

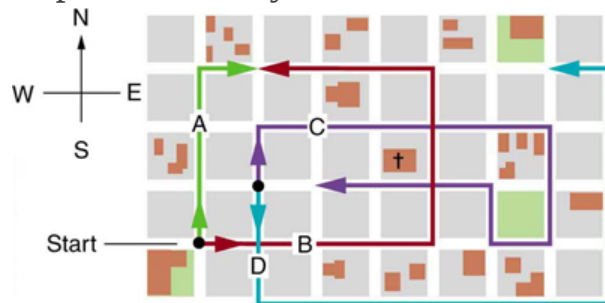
If the vectors **A** and **B** are perpendicular, what is the component of **A** along the direction of **B**? What is the component of **B** along the direction of **A**?

Problems & Exercises

Exercise:

Problem:

Find the following for path C in [\[link\]](#): (a) the total distance traveled and (b) the magnitude and direction of the displacement from start to finish. In this part of the problem, explicitly show how you follow the steps of the analytical method of vector addition.



The various lines represent paths taken by different people walking in a city. All blocks are 120 m on a side.

Solution:

(a) 1.56 km

(b) 120 m east

Exercise:**Problem:**

Find the following for path D in [\[link\]](#): (a) the total distance traveled and (b) the magnitude and direction of the displacement from start to finish. In this part of the problem, explicitly show how you follow the steps of the analytical method of vector addition.

Exercise:

Problem:

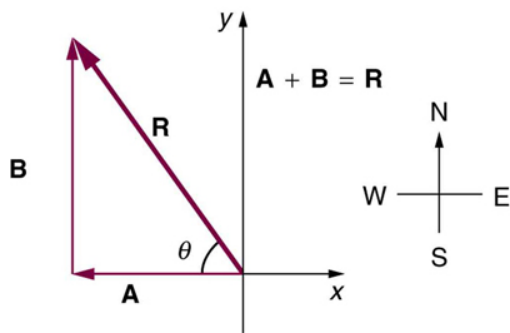
Find the north and east components of the displacement from San Francisco to Sacramento shown in [\[link\]](#).

**Solution:**

North-component 87.0 km, east-component 87.0 km

Exercise:**Problem:**

Solve the following problem using analytical techniques: Suppose you walk 18.0 m straight west and then 25.0 m straight north. How far are you from your starting point, and what is the compass direction of a line connecting your starting point to your final position? (If you represent the two legs of the walk as vector displacements **A** and **B**, as in [\[link\]](#), then this problem asks you to find their sum $\mathbf{R} = \mathbf{A} + \mathbf{B}$.)



The two displacements **A** and **B** add to give a total displacement **R** having magnitude R and direction θ .

Note that you can also solve this graphically. Discuss why the analytical technique for solving this problem is potentially more accurate than the graphical technique.

Exercise:

Problem:

Repeat [\[link\]](#) using analytical techniques, but reverse the order of the two legs of the walk and show that you get the same final result. (This problem shows that adding them in reverse order gives the same result—that is, $\mathbf{B} + \mathbf{A} = \mathbf{A} + \mathbf{B}$.) Discuss how taking another path to reach the same point might help to overcome an obstacle blocking your other path.

Solution:

30.8 m, 35.8 west of north

Exercise:

Problem:

You drive 7.50 km in a straight line in a direction 15° east of north. (a) Find the distances you would have to drive straight east and then straight north to arrive at the same point. (This determination is equivalent to find the components of the displacement along the east and north directions.) (b) Show that you still arrive at the same point if the east and north legs are reversed in order.

Exercise:**Problem:**

Do [\[link\]](#) again using analytical techniques and change the second leg of the walk to 25.0 m straight south. (This is equivalent to subtracting \mathbf{B} from \mathbf{A} —that is, finding $\mathbf{R} = \mathbf{A} - \mathbf{B}$) (b) Repeat again, but now you first walk 25.0 m north and then 18.0 m east. (This is equivalent to subtract \mathbf{A} from \mathbf{B} —that is, to find $\mathbf{A} = \mathbf{B} + \mathbf{C}$. Is that consistent with your result?)

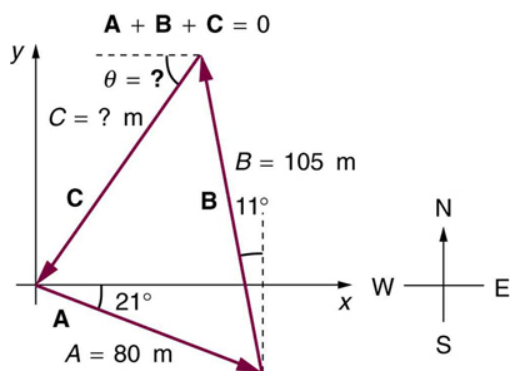
Solution:

(a) 30.8 m, 54.2° south of west

(b) 30.8 m, 54.2° north of east

Exercise:**Problem:**

A new landowner has a triangular piece of flat land she wishes to fence. Starting at the west corner, she measures the first side to be 80.0 m long and the next to be 105 m. These sides are represented as displacement vectors \mathbf{A} from \mathbf{B} in [\[link\]](#). She then correctly calculates the length and orientation of the third side \mathbf{C} . What is her result?



Exercise:

Problem:

You fly 32.0 km in a straight line in still air in the direction 35.0° south of west. (a) Find the distances you would have to fly straight south and then straight west to arrive at the same point. (This determination is equivalent to finding the components of the displacement along the south and west directions.) (b) Find the distances you would have to fly first in a direction 45.0° south of west and then in a direction 45.0° west of north. These are the components of the displacement along a different set of axes—one rotated 45° .

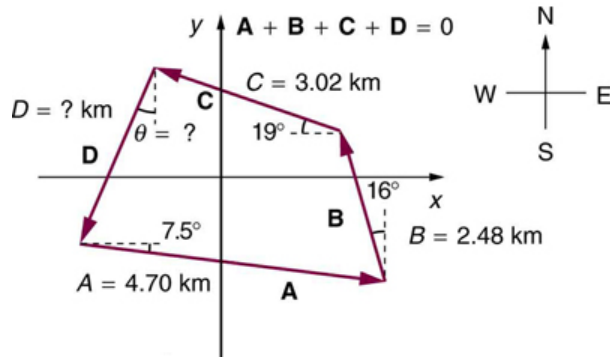
Solution:

18.4 km south, then 26.2 km west (b) 31.5 km at 45.0° south of west, then 5.56 km at 45.0° west of north

Exercise:

Problem:

A farmer wants to fence off his four-sided plot of flat land. He measures the first three sides, shown as **A**, **B**, and **C** in [\[link\]](#), and then correctly calculates the length and orientation of the fourth side **D**. What is his result?



Exercise:

Problem:

In an attempt to escape his island, Gilligan builds a raft and sets to sea. The wind shifts a great deal during the day, and he is blown along the following straight lines: 2.50 km 45.0° north of west; then 4.70 km 60.0° south of east; then 1.30 km 25.0° south of west; then 5.10 km straight east; then 1.70 km 5.00° east of north; then 7.20 km 55.0° south of west; and finally 2.80 km 10.0° north of east. What is his final position relative to the island?

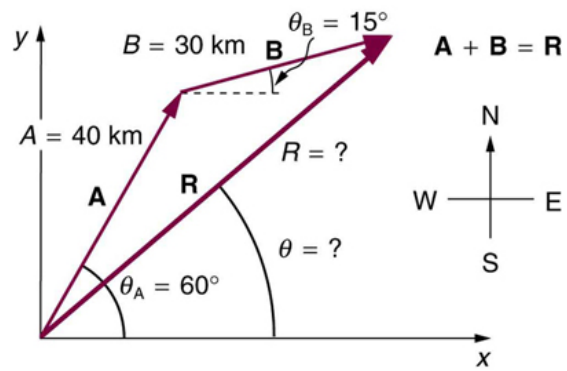
Solution:

7.34 km, 63.5° south of east

Exercise:

Problem:

Suppose a pilot flies 40.0 km in a direction 60° north of east and then flies 30.0 km in a direction 15° north of east as shown in [\[link\]](#). Find her total distance R from the starting point and the direction θ of the straight-line path to the final position. Discuss qualitatively how this flight would be altered by a wind from the north and how the effect of the wind would depend on both wind speed and the speed of the plane relative to the air mass.



Glossary

analytical method

the method of determining the magnitude and direction of a resultant vector using the Pythagorean theorem and trigonometric identities

Projectile Motion

- Identify and explain the properties of a projectile, such as acceleration due to gravity, range, maximum height, and trajectory.
- Determine the location and velocity of a projectile at different points in its trajectory.
- Apply the principle of independence of motion to solve projectile motion problems.

Projectile motion is the **motion** of an object thrown or projected into the air, subject to only the acceleration of gravity. The object is called a **projectile**, and its path is called its **trajectory**. The motion of falling objects, as covered in [Problem-Solving Basics for One-Dimensional Kinematics](#), is a simple one-dimensional type of projectile motion in which there is no horizontal movement. In this section, we consider two-dimensional projectile motion, such as that of a football or other object for which **air resistance is negligible**.

The most important fact to remember here is that *motions along perpendicular axes are independent* and thus can be analyzed separately. This fact was discussed in [Kinematics in Two Dimensions: An Introduction](#), where vertical and horizontal motions were seen to be independent. The key to analyzing two-dimensional projectile motion is to break it into two motions, one along the horizontal axis and the other along the vertical. (This choice of axes is the most sensible, because acceleration due to gravity is vertical—thus, there will be no acceleration along the horizontal axis when air resistance is negligible.) As is customary, we call the horizontal axis the x -axis and the vertical axis the y -axis. [\[link\]](#) illustrates the notation for displacement, where \mathbf{s} is defined to be the total displacement and \mathbf{x} and \mathbf{y} are its components along the horizontal and vertical axes, respectively. The magnitudes of these vectors are s , x , and y . (Note that in the last section we used the notation \mathbf{A} to represent a vector with components \mathbf{A}_x and \mathbf{A}_y . If we continued this format, we would call displacement \mathbf{s} with components \mathbf{s}_x and \mathbf{s}_y . However, to simplify the notation, we will simply represent the component vectors as \mathbf{x} and \mathbf{y} .)

Of course, to describe motion we must deal with velocity and acceleration, as well as with displacement. We must find their components along the x - and y -axes, too. We will assume all forces except gravity (such as air resistance and friction, for example) are negligible. The components of acceleration are then very simple: $a_y = -g = -9.80 \text{ m/s}^2$. (Note that this definition assumes that the upwards direction is defined as the positive direction. If you arrange the coordinate system instead such that the downwards direction is positive, then acceleration due to gravity takes a positive value.) Because gravity is vertical, $a_x = 0$. Both accelerations are constant, so the kinematic equations can be used.

Note:

Review of Kinematic Equations (constant a)

Equation:

$$x = x_0 + \bar{v}t$$

Equation:

$$\bar{v} = \frac{v_0 + v}{2}$$

Equation:

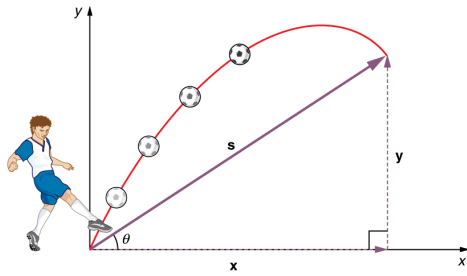
$$v = v_0 + at$$

Equation:

$$x = x_0 + v_0t + \frac{1}{2}at^2$$

Equation:

$$v^2 = v_0^2 + 2a(x - x_0).$$



The total displacement \mathbf{s} of a soccer ball at a point along its path. The vector \mathbf{s} has components \mathbf{x} and \mathbf{y} along the horizontal and vertical axes. Its magnitude is s , and it makes an angle θ with the horizontal.

Given these assumptions, the following steps are then used to analyze projectile motion:

Step 1. Resolve or break the motion into horizontal and vertical components along the x - and y -axes. These axes are perpendicular, so $A_x = A \cos \theta$ and $A_y = A \sin \theta$ are used. The magnitude of the components of displacement \mathbf{s} along these axes are x and y . The magnitudes of the components of the velocity \mathbf{v} are $v_x = v \cos \theta$ and $v_y = v \sin \theta$, where v is the magnitude of the velocity and θ is its direction, as shown in [\[link\]](#). Initial values are denoted with a subscript 0, as usual.

Step 2. Treat the motion as two independent one-dimensional motions, one horizontal and the other vertical. The kinematic equations for horizontal and vertical motion take the following forms:

Equation:

$$\text{Horizontal Motion}(a_x = 0)$$

Equation:

$$x = x_0 + v_x t$$

Equation:

$$v_x = v_{0x} = v_x = \text{velocity is a constant.}$$

Equation:

$$\text{Vertical Motion(assuming positive is up } a_y = -g = -9.80\text{m/s}^2)$$

Equation:

$$y = y_0 + \frac{1}{2}(v_{0y} + v_y)t$$

Equation:

$$v_y = v_{0y} - gt$$

Equation:

$$y = y_0 + v_{0y}t - \frac{1}{2}gt^2$$

Equation:

$$v_y^2 = v_{0y}^2 - 2g(y - y_0).$$

Step 3. Solve for the unknowns in the two separate motions—one horizontal and one vertical. Note that the only common variable between the motions is time t . The problem solving procedures here are the same as for one-dimensional **kinematics** and are illustrated in the solved examples below.

Step 4. Recombine the two motions to find the total displacement \mathbf{s} and velocity \mathbf{v} . Because the x - and y -motions are perpendicular, we determine these vectors by using the techniques outlined in the [Vector Addition and Subtraction: Analytical Methods](#) and employing $A = \sqrt{A_x^2 + A_y^2}$ and $\theta = \tan^{-1}(A_y/A_x)$ in the following form, where θ is the direction of the displacement \mathbf{s} and θ_v is the direction of the velocity \mathbf{v} :

Total displacement and velocity

Equation:

$$s = \sqrt{x^2 + y^2}$$

Equation:

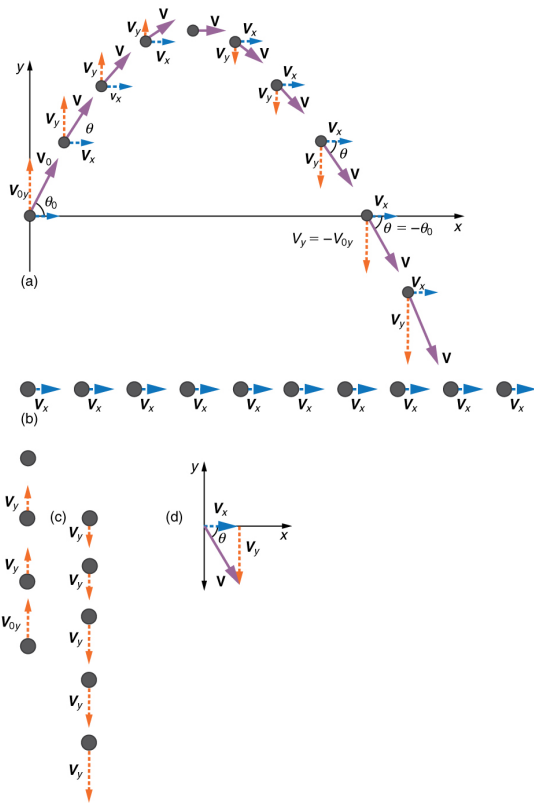
$$\theta = \tan^{-1}(y/x)$$

Equation:

$$v = \sqrt{v_x^2 + v_y^2}$$

Equation:

$$\theta_v = \tan^{-1}(v_y/v_x).$$



(a) We analyze two-dimensional projectile motion by breaking it into two independent one-dimensional motions along the vertical and horizontal axes. (b) The horizontal motion is simple, because $a_x = 0$ and v_x is thus constant. (c) The velocity in the vertical direction begins to decrease as the object rises; at its highest point, the vertical velocity is zero. As the object falls towards the Earth again, the vertical velocity increases again in magnitude but points in the opposite direction to the initial vertical velocity. (d) The x - and y -motions are recombined to give the total velocity at any given point on the trajectory.

Example:

A Fireworks Projectile Explodes High and Away

During a fireworks display, a shell is shot into the air with an initial speed of 70.0 m/s at an angle of 75.0° above the horizontal, as illustrated in [\[link\]](#). The fuse is timed to ignite the shell just as it reaches its highest point above the ground. (a) Calculate the height at which the shell explodes. (b) How much time passed between the launch of the shell and the explosion? (c) What is the horizontal displacement of the shell when it explodes?

Strategy

Because air resistance is negligible for the unexploded shell, the analysis method outlined above can be used. The motion can be broken into horizontal and vertical motions in which $a_x = 0$ and $a_y = -g$. We can then define x_0

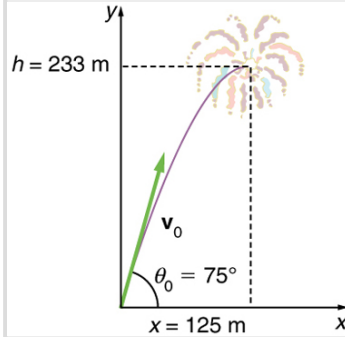
and y_0 to be zero and solve for the desired quantities.

Solution for (a)

By “height” we mean the altitude or vertical position y above the starting point. The highest point in any trajectory, called the apex, is reached when $v_y = 0$. Since we know the initial and final velocities as well as the initial position, we use the following equation to find y :

Equation:

$$v_y^2 = v_{0y}^2 - 2g(y - y_0).$$



The trajectory of a fireworks shell. The fuse is set to explode the shell at the highest point in its trajectory, which is found to be at a height of 233 m and 125 m away horizontally.

Because y_0 and v_y are both zero, the equation simplifies to

Equation:

$$0 = v_{0y}^2 - 2gy.$$

Solving for y gives

Equation:

$$y = \frac{v_{0y}^2}{2g}.$$

Now we must find v_{0y} , the component of the initial velocity in the y -direction. It is given by $v_{0y} = v_0 \sin \theta$, where v_0 is the initial velocity of 70.0 m/s, and $\theta = 75.0^\circ$ is the initial angle. Thus,

Equation:

$$v_{0y} = v_0 \sin \theta_0 = (70.0 \text{ m/s})(\sin 75^\circ) = 67.6 \text{ m/s}.$$

and y is

Equation:

$$y = \frac{(67.6 \text{ m/s})^2}{2(9.80 \text{ m/s}^2)},$$

so that

Equation:

$$y = 233\text{m}.$$

Discussion for (a)

Note that because up is positive, the initial velocity is positive, as is the maximum height, but the acceleration due to gravity is negative. Note also that the maximum height depends only on the vertical component of the initial velocity, so that any projectile with a 67.6 m/s initial vertical component of velocity will reach a maximum height of 233 m (neglecting air resistance). The numbers in this example are reasonable for large fireworks displays, the shells of which do reach such heights before exploding. In practice, air resistance is not completely negligible, and so the initial velocity would have to be somewhat larger than that given to reach the same height.

Solution for (b)

As in many physics problems, there is more than one way to solve for the time to the highest point. In this case, the easiest method is to use $y = y_0 + \frac{1}{2}(v_{0y} + v_y)t$. Because y_0 is zero, this equation reduces to simply

Equation:

$$y = \frac{1}{2}(v_{0y} + v_y)t.$$

Note that the final vertical velocity, v_y , at the highest point is zero. Thus,

Equation:

$$\begin{aligned} t &= \frac{2y}{(v_{0y} + v_y)} = \frac{2(233\text{ m})}{(67.6\text{ m/s})} \\ &= 6.90\text{ s}. \end{aligned}$$

Discussion for (b)

This time is also reasonable for large fireworks. When you are able to see the launch of fireworks, you will notice several seconds pass before the shell explodes. (Another way of finding the time is by using $y = y_0 + v_{0y}t - \frac{1}{2}gt^2$, and solving the quadratic equation for t .)

Solution for (c)

Because air resistance is negligible, $a_x = 0$ and the horizontal velocity is constant, as discussed above. The horizontal displacement is horizontal velocity multiplied by time as given by $x = x_0 + v_x t$, where x_0 is equal to zero:

Equation:

$$x = v_x t,$$

where v_x is the x-component of the velocity, which is given by $v_x = v_0 \cos \theta_0$. Now,

Equation:

$$v_x = v_0 \cos \theta_0 = (70.0\text{ m/s})(\cos 75.0^\circ) = 18.1\text{ m/s}.$$

The time t for both motions is the same, and so x is

Equation:

$$x = (18.1\text{ m/s})(6.90\text{ s}) = 125\text{ m}.$$

Discussion for (c)

The horizontal motion is a constant velocity in the absence of air resistance. The horizontal displacement found here could be useful in keeping the fireworks fragments from falling on spectators. Once the shell explodes, air resistance has a major effect, and many fragments will land directly below.

In solving part (a) of the preceding example, the expression we found for y is valid for any projectile motion where air resistance is negligible. Call the maximum height $y = h$; then,

Equation:

$$h = \frac{v_{0y}^2}{2g}.$$

This equation defines the *maximum height of a projectile* and depends only on the vertical component of the initial velocity.

Note:

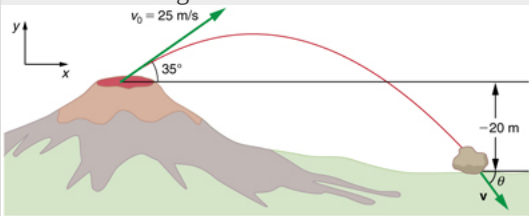
Defining a Coordinate System

It is important to set up a coordinate system when analyzing projectile motion. One part of defining the coordinate system is to define an origin for the x and y positions. Often, it is convenient to choose the initial position of the object as the origin such that $x_0 = 0$ and $y_0 = 0$. It is also important to define the positive and negative directions in the x and y directions. Typically, we define the positive vertical direction as upwards, and the positive horizontal direction is usually the direction of the object's motion. When this is the case, the vertical acceleration, g , takes a negative value (since it is directed downwards towards the Earth). However, it is occasionally useful to define the coordinates differently. For example, if you are analyzing the motion of a ball thrown downwards from the top of a cliff, it may make sense to define the positive direction downwards since the motion of the ball is solely in the downwards direction. If this is the case, g takes a positive value.

Example:

Calculating Projectile Motion: Hot Rock Projectile

Kilauea in Hawaii is the world's most continuously active volcano. Very active volcanoes characteristically eject red-hot rocks and lava rather than smoke and ash. Suppose a large rock is ejected from the volcano with a speed of 25.0 m/s and at an angle 35.0° above the horizontal, as shown in [link](#). The rock strikes the side of the volcano at an altitude 20.0 m lower than its starting point. (a) Calculate the time it takes the rock to follow this path. (b) What are the magnitude and direction of the rock's velocity at impact?



The trajectory of a rock ejected from the Kilauea volcano.

Strategy

Again, resolving this two-dimensional motion into two independent one-dimensional motions will allow us to solve for the desired quantities. The time a projectile is in the air is governed by its vertical motion alone. We will solve for t first. While the rock is rising and falling vertically, the horizontal motion continues at a constant velocity. This example asks for the final velocity. Thus, the vertical and horizontal results will be recombined to obtain v and θ_v at the final time t determined in the first part of the example.

Solution for (a)

While the rock is in the air, it rises and then falls to a final position 20.0 m lower than its starting altitude. We can find the time for this by using

Equation:

$$y = y_0 + v_{0y}t - \frac{1}{2}gt^2.$$

If we take the initial position y_0 to be zero, then the final position is $y = -20.0$ m. Now the initial vertical velocity is the vertical component of the initial velocity, found from $v_{0y} = v_0 \sin \theta_0 = (25.0 \text{ m/s})(\sin 35.0^\circ) = 14.3 \text{ m/s}$. Substituting known values yields

Equation:

$$-20.0 \text{ m} = (14.3 \text{ m/s})t - (4.90 \text{ m/s}^2)t^2.$$

Rearranging terms gives a quadratic equation in t :

Equation:

$$(4.90 \text{ m/s}^2)t^2 - (14.3 \text{ m/s})t - (20.0 \text{ m}) = 0.$$

This expression is a quadratic equation of the form $at^2 + bt + c = 0$, where the constants are $a = 4.90$, $b = -14.3$, and $c = -20.0$. Its solutions are given by the quadratic formula:

Equation:

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

This equation yields two solutions: $t = 3.96$ and $t = -1.03$. (It is left as an exercise for the reader to verify these solutions.) The time is $t = 3.96$ s or -1.03 s. The negative value of time implies an event before the start of motion, and so we discard it. Thus,

Equation:

$$t = 3.96 \text{ s}.$$

Discussion for (a)

The time for projectile motion is completely determined by the vertical motion. So any projectile that has an initial vertical velocity of 14.3 m/s and lands 20.0 m below its starting altitude will spend 3.96 s in the air.

Solution for (b)

From the information now in hand, we can find the final horizontal and vertical velocities v_x and v_y and combine them to find the total velocity v and the angle θ_0 it makes with the horizontal. Of course, v_x is constant so we can solve for it at any horizontal location. In this case, we chose the starting point since we know both the initial velocity and initial angle. Therefore:

Equation:

$$v_x = v_0 \cos \theta_0 = (25.0 \text{ m/s})(\cos 35^\circ) = 20.5 \text{ m/s}.$$

The final vertical velocity is given by the following equation:

Equation:

$$v_y = v_{0y} - gt,$$

where v_{0y} was found in part (a) to be 14.3 m/s. Thus,

Equation:

$$v_y = 14.3 \text{ m/s} - (9.80 \text{ m/s}^2)(3.96 \text{ s})$$

so that

Equation:

$$v_y = -24.5 \text{ m/s}.$$

To find the magnitude of the final velocity v we combine its perpendicular components, using the following equation:

Equation:

$$v = \sqrt{v_x^2 + v_y^2} = \sqrt{(20.5 \text{ m/s})^2 + (-24.5 \text{ m/s})^2},$$

which gives

Equation:

$$v = 31.9 \text{ m/s}.$$

The direction θ_v is found from the equation:

Equation:

$$\theta_v = \tan^{-1}(v_y/v_x)$$

so that

Equation:

$$\theta_v = \tan^{-1}(-24.5/20.5) = \tan^{-1}(-1.19).$$

Thus,

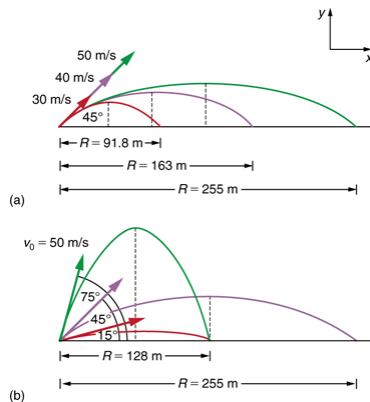
Equation:

$$\theta_v = -50.1^\circ.$$

Discussion for (b)

The negative angle means that the velocity is 50.1° below the horizontal. This result is consistent with the fact that the final vertical velocity is negative and hence downward—as you would expect because the final altitude is 20.0 m lower than the initial altitude. (See [\[link\]](#).)

One of the most important things illustrated by projectile motion is that vertical and horizontal motions are independent of each other. Galileo was the first person to fully comprehend this characteristic. He used it to predict the range of a projectile. On level ground, we define **range** to be the horizontal distance R traveled by a projectile. Galileo and many others were interested in the range of projectiles primarily for military purposes—such as aiming cannons. However, investigating the range of projectiles can shed light on other interesting phenomena, such as the orbits of satellites around the Earth. Let us consider projectile range further.



Trajectories of projectiles on level ground. (a) The greater the initial speed v_0 , the greater the range for a given initial angle. (b) The effect of initial angle θ_0 on the range of a projectile with a

given initial speed. Note that the range is the same for 15° and 75° , although the maximum heights of those paths are different.

How does the initial velocity of a projectile affect its range? Obviously, the greater the initial speed v_0 , the greater the range, as shown in [\[link\]](#)(a). The initial angle θ_0 also has a dramatic effect on the range, as illustrated in [\[link\]](#)(b). For a fixed initial speed, such as might be produced by a cannon, the maximum range is obtained with $\theta_0 = 45^\circ$. This is true only for conditions neglecting air resistance. If air resistance is considered, the maximum angle is approximately 38° . Interestingly, for every initial angle except 45° , there are two angles that give the same range—the sum of those angles is 90° . The range also depends on the value of the acceleration of gravity g . The lunar astronaut Alan Shepherd was able to drive a golf ball a great distance on the Moon because gravity is weaker there. The range R of a projectile on *level ground* for which air resistance is negligible is given by

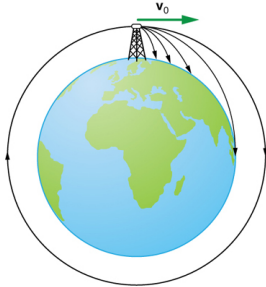
Equation:

$$R = \frac{v_0^2 \sin 2\theta_0}{g},$$

where v_0 is the initial speed and θ_0 is the initial angle relative to the horizontal. The proof of this equation is left as an end-of-chapter problem (hints are given), but it does fit the major features of projectile range as described.

When we speak of the range of a projectile on level ground, we assume that R is very small compared with the circumference of the Earth. If, however, the range is large, the Earth curves away below the projectile and acceleration of gravity changes direction along the path. The range is larger than predicted by the range equation given above because the projectile has farther to fall than it would on level ground. (See [\[link\]](#).) If the initial speed is great enough, the projectile goes into orbit. This possibility was recognized centuries before it could be accomplished. When an object is in orbit, the Earth curves away from underneath the object at the same rate as it falls. The object thus falls continuously but never hits the surface. These and other aspects of orbital motion, such as the rotation of the Earth, will be covered analytically and in greater depth later in this text.

Once again we see that thinking about one topic, such as the range of a projectile, can lead us to others, such as the Earth orbits. In [Addition of Velocities](#), we will examine the addition of velocities, which is another important aspect of two-dimensional kinematics and will also yield insights beyond the immediate topic.



Projectile to satellite. In each case shown here, a projectile is launched from a very high tower to avoid air resistance. With increasing initial speed, the

range increases and becomes longer than it would be on level ground because the Earth curves away underneath its path. With a large enough initial speed, orbit is achieved.

Note:

PhET Explorations: Projectile Motion

Blast a Buick out of a cannon! Learn about projectile motion by firing various objects. Set the angle, initial speed, and mass. Add air resistance. Make a game out of this simulation by trying to hit a target.

https://phet.colorado.edu/sims/projectile-motion/projectile-motion_en.html

Summary

- Projectile motion is the motion of an object through the air that is subject only to the acceleration of gravity.
- To solve projectile motion problems, perform the following steps:

Determine a coordinate system. Then, resolve the position and/or velocity of the object in the horizontal and vertical components. The components of position are given by the quantities x and y , and the components of the velocity \mathbf{v} are $v_x = v \cos \theta$ and $v_y = v \sin \theta$, where v is the magnitude of the velocity and θ is its direction.

Analyze the motion of the projectile in the horizontal direction using the following equations:

Equation:

Horizontal motion ($a_x = 0$)

Equation:

$$x = x_0 + v_x t$$

Equation:

$v_x = v_{0x} = \mathbf{v}_x = \text{velocity is a constant.}$

Analyze the motion of the projectile in the vertical

Equation:

Equation:

Vertical motion (Assuming positive direction is up; $a_y = -g = -9.80 \text{ m/s}^2$)
 $y = y_0 + \frac{1}{2}(v_{0y} + v_y)t$

direction
using the
following
equations:

Recombine the
horizontal and
vertical components
of location and/or
velocity using the
following equations:

Equation:

$$s = \sqrt{x^2 + y^2}$$

Equation:

$$\theta = \tan^{-1}(y/x)$$

Equation:

$$v = \sqrt{v_x^2 + v_y^2}$$

Equation:

$$\theta_v = \tan^{-1}(v_y/v_x).$$

- The maximum height h of a projectile launched with initial vertical velocity v_{0y} is given by

Equation:

$$h = \frac{v_{0y}^2}{2g}.$$

- The maximum horizontal distance traveled by a projectile is called the **range**. The range R of a projectile on level ground launched at an angle θ_0 above the horizontal with initial speed v_0 is given by

Equation:

$$R = \frac{v_0^2 \sin 2\theta_0}{g}.$$

Conceptual Questions

Exercise:

Problem:

Answer the following questions for projectile motion on level ground assuming negligible air resistance (the initial angle being neither 0° nor 90°): (a) Is the velocity ever zero? (b) When is the velocity a minimum? A maximum? (c) Can the velocity ever be the same as the initial velocity at a time other than at $t = 0$? (d) Can the speed ever be the same as the initial speed at a time other than at $t = 0$?

Exercise:

Problem:

Answer the following questions for projectile motion on level ground assuming negligible air resistance (the initial angle being neither 0° nor 90°): (a) Is the acceleration ever zero? (b) Is the acceleration ever in the same direction as a component of velocity? (c) Is the acceleration ever opposite in direction to a component of velocity?

Exercise:

Problem:

For a fixed initial speed, the range of a projectile is determined by the angle at which it is fired. For all but the maximum, there are two angles that give the same range. Considering factors that might affect the ability of an archer to hit a target, such as wind, explain why the smaller angle (closer to the horizontal) is preferable. When would it be necessary for the archer to use the larger angle? Why does the punter in a football game use the higher trajectory?

Exercise:

Problem:

During a lecture demonstration, a professor places two coins on the edge of a table. She then flicks one of the coins horizontally off the table, simultaneously nudging the other over the edge. Describe the subsequent motion of the two coins, in particular discussing whether they hit the floor at the same time.

Problems & Exercises**Exercise:****Problem:**

A projectile is launched at ground level with an initial speed of 50.0 m/s at an angle of 30.0° above the horizontal. It strikes a target above the ground 3.00 seconds later. What are the x and y distances from where the projectile was launched to where it lands?

Solution:

$$x = 1.30 \text{ m} \times 10^2$$

$$y = 30.9 \text{ m}.$$

Exercise:**Problem:**

A ball is kicked with an initial velocity of 16 m/s in the horizontal direction and 12 m/s in the vertical direction. (a) At what speed does the ball hit the ground? (b) For how long does the ball remain in the air? (c) What maximum height is attained by the ball?

Exercise:**Problem:**

A ball is thrown horizontally from the top of a 60.0-m building and lands 100.0 m from the base of the building. Ignore air resistance. (a) How long is the ball in the air? (b) What must have been the initial horizontal component of the velocity? (c) What is the vertical component of the velocity just before the ball hits the ground? (d) What is the velocity (including both the horizontal and vertical components) of the ball just before it hits the ground?

Solution:

(a) 3.50 s

(b) 28.6 m/s (c) 34.3 m/s

(d) 44.7 m/s, 50.2° below horizontal

Exercise:**Problem:**

(a) A daredevil is attempting to jump his motorcycle over a line of buses parked end to end by driving up a 32° ramp at a speed of 40.0 m/s (144 km/h). How many buses can he clear if the top of the takeoff ramp is at the same height as the bus tops and the buses are 20.0 m long? (b) Discuss what your answer implies about the margin of error in this act—that is, consider how much greater the range is than the horizontal distance he must travel to miss the end of the last bus. (Neglect air resistance.)

Exercise:

Problem:

An archer shoots an arrow at a 75.0 m distant target; the bull's-eye of the target is at same height as the release height of the arrow. (a) At what angle must the arrow be released to hit the bull's-eye if its initial speed is 35.0 m/s? In this part of the problem, explicitly show how you follow the steps involved in solving projectile motion problems. (b) There is a large tree halfway between the archer and the target with an overhanging horizontal branch 3.50 m above the release height of the arrow. Will the arrow go over or under the branch?

Solution:

(a) 18.4°

(b) The arrow will go over the branch.

Exercise:**Problem:**

A rugby player passes the ball 7.00 m across the field, where it is caught at the same height as it left his hand. (a) At what angle was the ball thrown if its initial speed was 12.0 m/s, assuming that the smaller of the two possible angles was used? (b) What other angle gives the same range, and why would it not be used? (c) How long did this pass take?

Exercise:

Problem: Verify the ranges for the projectiles in [\[link\]](#)(a) for $\theta = 45^\circ$ and the given initial velocities.

Solution:

$$R = \frac{v_0^2}{\sin 2\theta_0 g}$$

$$\text{For } \theta = 45^\circ, \quad R = \frac{v_0^2}{g}$$

$$R = 91.8 \text{ m for } v_0 = 30 \text{ m/s}; \quad R = 163 \text{ m for } v_0 = 40 \text{ m/s}; \quad R = 255 \text{ m for } v_0 = 50 \text{ m/s}.$$

Exercise:**Problem:**

Verify the ranges shown for the projectiles in [\[link\]](#)(b) for an initial velocity of 50 m/s at the given initial angles.

Exercise:**Problem:**

The cannon on a battleship can fire a shell a maximum distance of 32.0 km. (a) Calculate the initial velocity of the shell. (b) What maximum height does it reach? (At its highest, the shell is above 60% of the atmosphere—but air resistance is not really negligible as assumed to make this problem easier.) (c) The ocean is not flat, because the Earth is curved. Assume that the radius of the Earth is 6.37×10^3 km. How many meters lower will its surface be 32.0 km from the ship along a horizontal line parallel to the surface at the ship? Does your answer imply that error introduced by the assumption of a flat Earth in projectile motion is significant here?

Solution:

(a) 560 m/s

(b) 8.00×10^3 m

(c) 80.0 m. This error is not significant because it is only 1% of the answer in part (b).

Exercise:

Problem:

An arrow is shot from a height of 1.5 m toward a cliff of height H . It is shot with a velocity of 30 m/s at an angle of 60° above the horizontal. It lands on the top edge of the cliff 4.0 s later. (a) What is the height of the cliff? (b) What is the maximum height reached by the arrow along its trajectory? (c) What is the arrow's impact speed just before hitting the cliff?

Exercise:

Problem:

In the standing broad jump, one squats and then pushes off with the legs to see how far one can jump. Suppose the extension of the legs from the crouch position is 0.600 m and the acceleration achieved from this position is 1.25 times the acceleration due to gravity, g . How far can they jump? State your assumptions. (Increased range can be achieved by swinging the arms in the direction of the jump.)

Solution:

1.50 m, assuming launch angle of 45°

Exercise:

Problem:

The world long jump record is 8.95 m (Mike Powell, USA, 1991). Treated as a projectile, what is the maximum range obtainable by a person if he has a take-off speed of 9.5 m/s? State your assumptions.

Exercise:

Problem:

Serving at a speed of 170 km/h, a tennis player hits the ball at a height of 2.5 m and an angle θ below the horizontal. The base line is 11.9 m from the net, which is 0.91 m high. What is the angle θ such that the ball just crosses the net? Will the ball land in the service box, whose service line is 6.40 m from the net?

Solution:

$$\theta = 6.1^\circ$$

yes, the ball lands at 5.3 m from the net

Exercise:

Problem:

A football quarterback is moving straight backward at a speed of 2.00 m/s when he throws a pass to a player 18.0 m straight downfield. (a) If the ball is thrown at an angle of 25° relative to the ground and is caught at the same height as it is released, what is its initial speed relative to the ground? (b) How long does it take to get to the receiver? (c) What is its maximum height above its point of release?

Exercise:

Problem:

Gun sights are adjusted to aim high to compensate for the effect of gravity, effectively making the gun accurate only for a specific range. (a) If a gun is sighted to hit targets that are at the same height as the gun and 100.0 m away, how low will the bullet hit if aimed directly at a target 150.0 m away? The muzzle velocity of the bullet is 275 m/s. (b) Discuss qualitatively how a larger muzzle velocity would affect this problem and what would be the effect of air resistance.

Solution:

(a) -0.486 m

(b) The larger the muzzle velocity, the smaller the deviation in the vertical direction, because the time of flight would be smaller. Air resistance would have the effect of decreasing the time of flight, therefore increasing the vertical deviation.

Exercise:**Problem:**

An eagle is flying horizontally at a speed of 3.00 m/s when the fish in her talons wiggles loose and falls into the lake 5.00 m below. Calculate the velocity of the fish relative to the water when it hits the water.

Exercise:**Problem:**

An owl is carrying a mouse to the chicks in its nest. Its position at that time is 4.00 m west and 12.0 m above the center of the 30.0 cm diameter nest. The owl is flying east at 3.50 m/s at an angle 30.0° below the horizontal when it accidentally drops the mouse. Is the owl lucky enough to have the mouse hit the nest? To answer this question, calculate the horizontal position of the mouse when it has fallen 12.0 m .

Solution:

4.23 m . No, the owl is not lucky; he misses the nest.

Exercise:**Problem:**

Suppose a soccer player kicks the ball from a distance 30 m toward the goal. Find the initial speed of the ball if it just passes over the goal, 2.4 m above the ground, given the initial direction to be 40° above the horizontal.

Exercise:**Problem:**

Can a goalkeeper at her/ his goal kick a soccer ball into the opponent's goal without the ball touching the ground? The distance will be about 95 m . A goalkeeper can give the ball a speed of 30 m/s .

Solution:

No, the maximum range (neglecting air resistance) is about 92 m .

Exercise:**Problem:**

The free throw line in basketball is 4.57 m (15 ft) from the basket, which is 3.05 m (10 ft) above the floor. A player standing on the free throw line throws the ball with an initial speed of 8.15 m/s , releasing it at a height of 2.44 m (8 ft) above the floor. At what angle above the horizontal must the ball be thrown to exactly hit the basket? Note that most players will use a large initial angle rather than a flat shot because it allows for a larger margin of error. Explicitly show how you follow the steps involved in solving projectile motion problems.

Exercise:

Problem:

In 2007, Michael Carter (U.S.) set a world record in the shot put with a throw of 24.77 m. What was the initial speed of the shot if he released it at a height of 2.10 m and threw it at an angle of 38.0° above the horizontal? (Although the maximum distance for a projectile on level ground is achieved at 45° when air resistance is neglected, the actual angle to achieve maximum range is smaller; thus, 38° will give a longer range than 45° in the shot put.)

Solution:

15.0 m/s

Exercise:**Problem:**

A basketball player is running at 5.00 m/s directly toward the basket when he jumps into the air to dunk the ball. He maintains his horizontal velocity. (a) What vertical velocity does he need to rise 0.750 m above the floor? (b) How far from the basket (measured in the horizontal direction) must he start his jump to reach his maximum height at the same time as he reaches the basket?

Exercise:**Problem:**

A football player punts the ball at a 45.0° angle. Without an effect from the wind, the ball would travel 60.0 m horizontally. (a) What is the initial speed of the ball? (b) When the ball is near its maximum height it experiences a brief gust of wind that reduces its horizontal velocity by 1.50 m/s. What distance does the ball travel horizontally?

Solution:

(a) 24.2 m/s

(b) The ball travels a total of 57.4 m with the brief gust of wind.

Exercise:**Problem:**

Prove that the trajectory of a projectile is parabolic, having the form $y = ax + bx^2$. To obtain this expression, solve the equation $x = v_{0x}t$ for t and substitute it into the expression for $y = v_{0y}t - (1/2)gt^2$ (These equations describe the x and y positions of a projectile that starts at the origin.) You should obtain an equation of the form $y = ax + bx^2$ where a and b are constants.

Exercise:**Problem:**

Derive $R = \frac{v_0^2 \sin 2\theta_0}{g}$ for the range of a projectile on level ground by finding the time t at which y becomes zero and substituting this value of t into the expression for $x - x_0$, noting that $R = x - x_0$

Solution:

$$y - y_0 = 0 = v_{0y}t - \frac{1}{2}gt^2 = (v_0 \sin \theta)t - \frac{1}{2}gt^2,$$

$$\text{so that } t = \frac{2(v_0 \sin \theta)}{g}$$

$x - x_0 = v_{0x}t = (v_0 \cos \theta)t = R$, and substituting for t gives:

$$R = v_0 \cos \theta \left(\frac{2v_0 \sin \theta}{g} \right) = \frac{2v_0^2 \sin \theta \cos \theta}{g}$$

since $2 \sin \theta \cos \theta = \sin 2\theta$, the range is:

$$R = \frac{v_0^2 \sin 2\theta}{g}.$$

Exercise:

Problem:

Unreasonable Results (a) Find the maximum range of a super cannon that has a muzzle velocity of 4.0 km/s. (b) What is unreasonable about the range you found? (c) Is the premise unreasonable or is the available equation inapplicable? Explain your answer. (d) If such a muzzle velocity could be obtained, discuss the effects of air resistance, thinning air with altitude, and the curvature of the Earth on the range of the super cannon.

Exercise:

Problem:

Construct Your Own Problem Consider a ball tossed over a fence. Construct a problem in which you calculate the ball's needed initial velocity to just clear the fence. Among the things to determine are; the height of the fence, the distance to the fence from the point of release of the ball, and the height at which the ball is released. You should also consider whether it is possible to choose the initial speed for the ball and just calculate the angle at which it is thrown. Also examine the possibility of multiple solutions given the distances and heights you have chosen.

Glossary

air resistance

a frictional force that slows the motion of objects as they travel through the air; when solving basic physics problems, air resistance is assumed to be zero

kinematics

the study of motion without regard to mass or force

motion

displacement of an object as a function of time

projectile

an object that travels through the air and experiences only acceleration due to gravity

projectile motion

the motion of an object that is subject only to the acceleration of gravity

range

the maximum horizontal distance that a projectile travels

trajectory

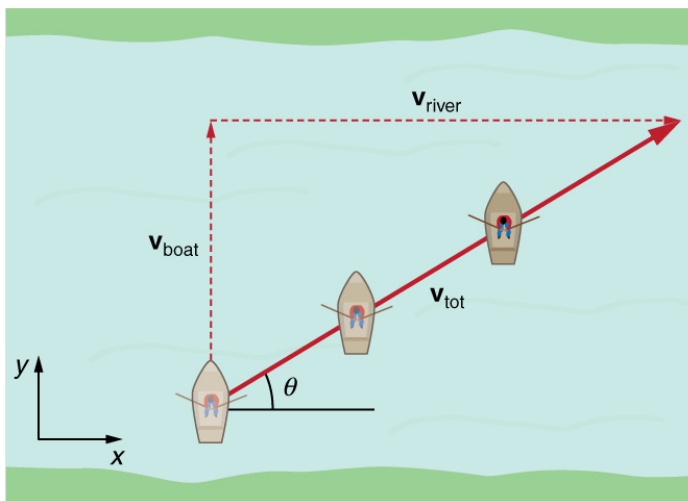
the path of a projectile through the air

Addition of Velocities

- Apply principles of vector addition to determine relative velocity.
- Explain the significance of the observer in the measurement of velocity.

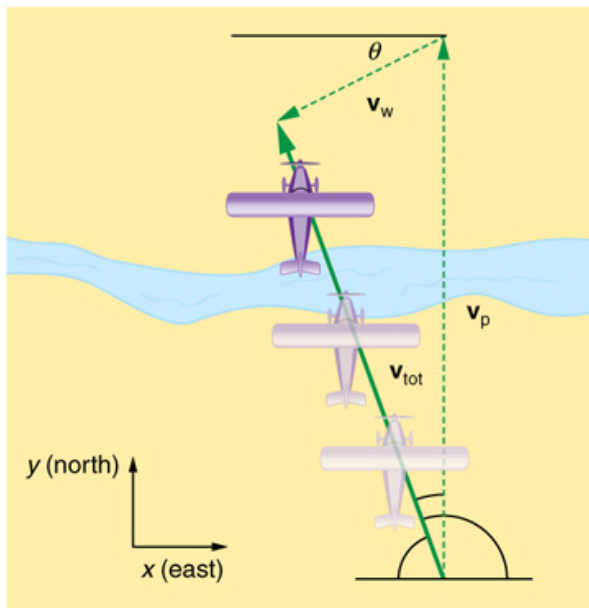
Relative Velocity

If a person rows a boat across a rapidly flowing river and tries to head directly for the other shore, the boat instead moves *diagonally* relative to the shore, as in [\[link\]](#). The boat does not move in the direction in which it is pointed. The reason, of course, is that the river carries the boat downstream. Similarly, if a small airplane flies overhead in a strong crosswind, you can sometimes see that the plane is not moving in the direction in which it is pointed, as illustrated in [\[link\]](#). The plane is moving straight ahead relative to the air, but the movement of the air mass relative to the ground carries it sideways.



A boat trying to head straight across a river will actually move diagonally relative to the shore as shown. Its total velocity (solid arrow) relative to the shore is the sum of its velocity relative

to the river plus the velocity of the river relative to the shore.



An airplane heading straight north is instead carried to the west and slowed down by wind.

The plane does not move relative to the ground in the direction it points; rather, it moves in the direction of its total velocity (solid arrow).

In each of these situations, an object has a **velocity** relative to a medium (such as a river) and that medium has a velocity relative to an observer on solid ground. The velocity of the object *relative to the observer* is the sum of these velocity vectors, as indicated in [\[link\]](#) and [\[link\]](#). These situations are only two of many in which it is useful to add velocities. In this module,

we first re-examine how to add velocities and then consider certain aspects of what relative velocity means.

How do we add velocities? Velocity is a vector (it has both magnitude and direction); the rules of **vector addition** discussed in [Vector Addition and Subtraction: Graphical Methods](#) and [Vector Addition and Subtraction: Analytical Methods](#) apply to the addition of velocities, just as they do for any other vectors. In one-dimensional motion, the addition of velocities is simple—they add like ordinary numbers. For example, if a field hockey player is moving at 5 m/s straight toward the goal and drives the ball in the same direction with a velocity of 30 m/s relative to her body, then the velocity of the ball is 35 m/s relative to the stationary, profusely sweating goalkeeper standing in front of the goal.

In two-dimensional motion, either graphical or analytical techniques can be used to add velocities. We will concentrate on analytical techniques. The following equations give the relationships between the magnitude and direction of velocity (v and θ) and its components (v_x and v_y) along the x - and y -axes of an appropriately chosen coordinate system:

Equation:

$$v_x = v \cos \theta$$

Equation:

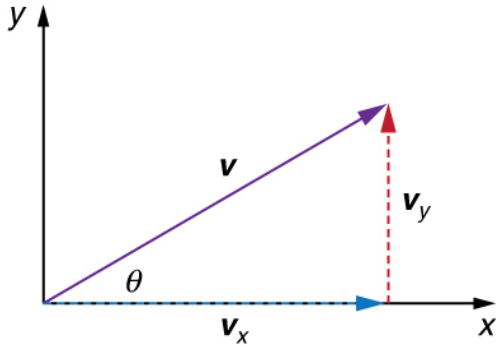
$$v_y = v \sin \theta$$

Equation:

$$v = \sqrt{v_x^2 + v_y^2}$$

Equation:

$$\theta = \tan^{-1}(v_y/v_x).$$



The velocity, v , of an object traveling at an angle θ to the horizontal axis is the sum of component vectors \mathbf{v}_x and \mathbf{v}_y .

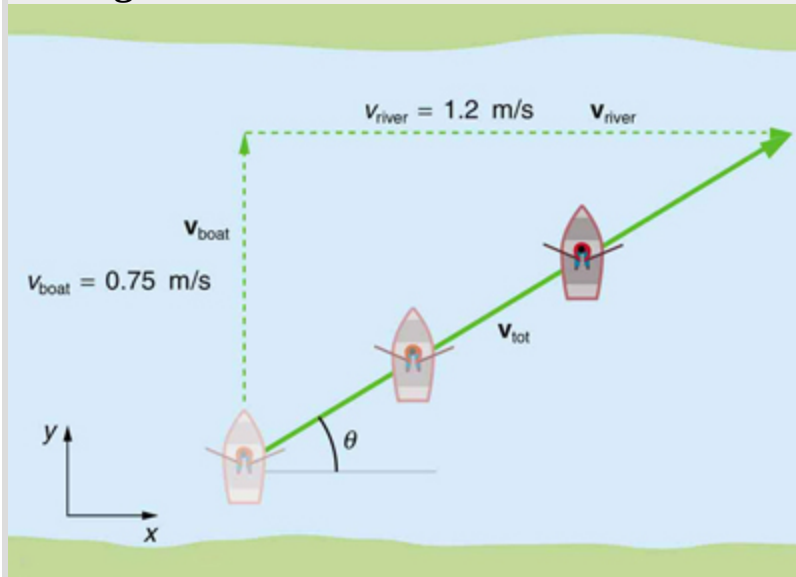
These equations are valid for any vectors and are adapted specifically for velocity. The first two equations are used to find the components of a velocity when its magnitude and direction are known. The last two are used to find the magnitude and direction of velocity when its components are known.

Note:**Take-Home Experiment: Relative Velocity of a Boat**

Fill a bathtub half-full of water. Take a toy boat or some other object that floats in water. Unplug the drain so water starts to drain. Try pushing the boat from one side of the tub to the other and perpendicular to the flow of water. Which way do you need to push the boat so that it ends up immediately opposite? Compare the directions of the flow of water, heading of the boat, and actual velocity of the boat.

Example:

Adding Velocities: A Boat on a River



A boat attempts to travel straight across a river at a speed 0.75 m/s . The current in the river, however, flows at a speed of 1.20 m/s to the right. What is the total displacement of the boat relative to the shore?

Refer to [\[link\]](#), which shows a boat trying to go straight across the river. Let us calculate the magnitude and direction of the boat's velocity relative to an observer on the shore, \mathbf{v}_{tot} . The velocity of the boat, \mathbf{v}_{boat} , is 0.75 m/s in the y -direction relative to the river and the velocity of the river, $\mathbf{v}_{\text{river}}$, is 1.20 m/s to the right.

Strategy

We start by choosing a coordinate system with its x -axis parallel to the velocity of the river, as shown in [\[link\]](#). Because the boat is directed straight toward the other shore, its velocity relative to the water is parallel to the y -axis and perpendicular to the velocity of the river. Thus, we can add the two velocities by using the equations $v_{\text{tot}} = \sqrt{v_x^2 + v_y^2}$ and $\theta = \tan^{-1}(v_y/v_x)$ directly.

Solution

The magnitude of the total velocity is

Equation:

$$v_{\text{tot}} = \sqrt{v_x^2 + v_y^2},$$

where

Equation:

$$v_x = v_{\text{river}} = 1.20 \text{ m/s}$$

and

Equation:

$$v_y = v_{\text{boat}} = 0.750 \text{ m/s}.$$

Thus,

Equation:

$$v_{\text{tot}} = \sqrt{(1.20 \text{ m/s})^2 + (0.750 \text{ m/s})^2}$$

yielding

Equation:

$$v_{\text{tot}} = 1.42 \text{ m/s}.$$

The direction of the total velocity θ is given by:

Equation:

$$\theta = \tan^{-1}(v_y/v_x) = \tan^{-1}(0.750/1.20).$$

This equation gives

Equation:

$$\theta = 32.0^\circ.$$

Discussion

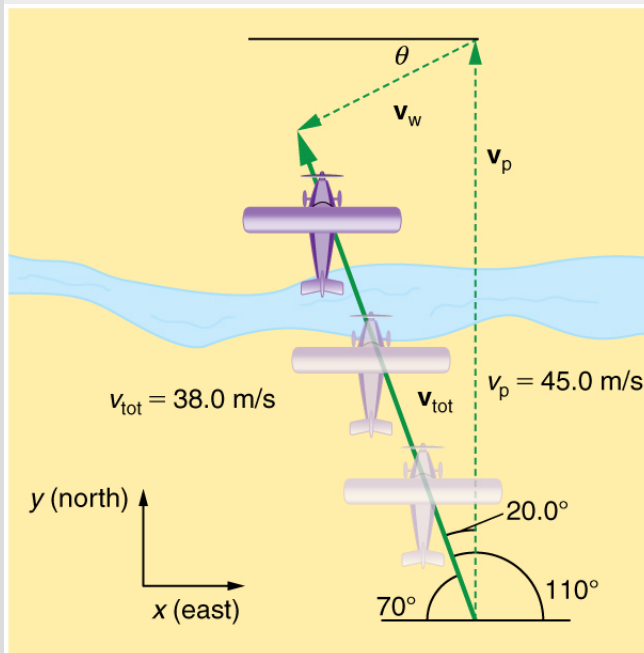
Both the magnitude v and the direction θ of the total velocity are consistent with [\[link\]](#). Note that because the velocity of the river is large compared

with the velocity of the boat, it is swept rapidly downstream. This result is evidenced by the small angle (only 32.0°) the total velocity has relative to the riverbank.

Example:

Calculating Velocity: Wind Velocity Causes an Airplane to Drift

Calculate the wind velocity for the situation shown in [\[link\]](#). The plane is known to be moving at 45.0 m/s due north relative to the air mass, while its velocity relative to the ground (its total velocity) is 38.0 m/s in a direction 20.0° west of north.



An airplane is known to be heading north at 45.0 m/s , though its velocity relative to the ground is 38.0 m/s at an angle west of north. What is the speed and direction of the wind?

Strategy

In this problem, somewhat different from the previous example, we know the total velocity \mathbf{v}_{tot} and that it is the sum of two other velocities, \mathbf{v}_w (the wind) and \mathbf{v}_p (the plane relative to the air mass). The quantity \mathbf{v}_p is known, and we are asked to find \mathbf{v}_w . None of the velocities are perpendicular, but it is possible to find their components along a common set of perpendicular axes. If we can find the components of \mathbf{v}_w , then we can combine them to solve for its magnitude and direction. As shown in [\[link\]](#), we choose a coordinate system with its x-axis due east and its y-axis due north (parallel to \mathbf{v}_p). (You may wish to look back at the discussion of the addition of vectors using perpendicular components in [Vector Addition and Subtraction: Analytical Methods](#).)

Solution

Because \mathbf{v}_{tot} is the vector sum of the \mathbf{v}_w and \mathbf{v}_p , its x- and y-components are the sums of the x- and y-components of the wind and plane velocities. Note that the plane only has vertical component of velocity so $v_{px} = 0$ and $v_{py} = v_p$. That is,

Equation:

$$v_{\text{tot}x} = v_{wx}$$

and

Equation:

$$v_{\text{tot}y} = v_{wy} + v_p.$$

We can use the first of these two equations to find v_{wx} :

Equation:

$$v_{wx} = v_{\text{tot}x} = v_{\text{tot}} \cos 110^\circ.$$

Because $v_{\text{tot}} = 38.0 \text{ m/s}$ and $\cos 110^\circ = -0.342$ we have

Equation:

$$v_{wx} = (38.0 \text{ m/s})(-0.342) = -13 \text{ m/s}.$$

The minus sign indicates motion west which is consistent with the diagram.

Now, to find v_{wy} we note that

Equation:

$$v_{\text{tot}y} = v_{\text{wy}} + v_{\text{p}}$$

Here $v_{\text{tot}y} = v_{\text{tot}} \sin 110^\circ$; thus,

Equation:

$$v_{\text{wy}} = (38.0 \text{ m/s})(0.940) - 45.0 \text{ m/s} = -9.29 \text{ m/s}.$$

This minus sign indicates motion south which is consistent with the diagram.

Now that the perpendicular components of the wind velocity v_{wx} and v_{wy} are known, we can find the magnitude and direction of \mathbf{v}_{w} . First, the magnitude is

Equation:

$$\begin{aligned} v_{\text{w}} &= \sqrt{v_{\text{wx}}^2 + v_{\text{wy}}^2} \\ &= \sqrt{(-13.0 \text{ m/s})^2 + (-9.29 \text{ m/s})^2} \end{aligned}$$

so that

Equation:

$$v_{\text{w}} = 16.0 \text{ m/s}.$$

The direction is:

Equation:

$$\theta = \tan^{-1}(v_{\text{wy}}/v_{\text{wx}}) = \tan^{-1}(-9.29/-13.0)$$

giving

Equation:

$$\theta = 35.6^\circ.$$

Discussion

The wind's speed and direction are consistent with the significant effect the wind has on the total velocity of the plane, as seen in [\[link\]](#). Because the

plane is fighting a strong combination of crosswind and head-wind, it ends up with a total velocity significantly less than its velocity relative to the air mass as well as heading in a different direction.

Note that in both of the last two examples, we were able to make the mathematics easier by choosing a coordinate system with one axis parallel to one of the velocities. We will repeatedly find that choosing an appropriate coordinate system makes problem solving easier. For example, in projectile motion we always use a coordinate system with one axis parallel to gravity.

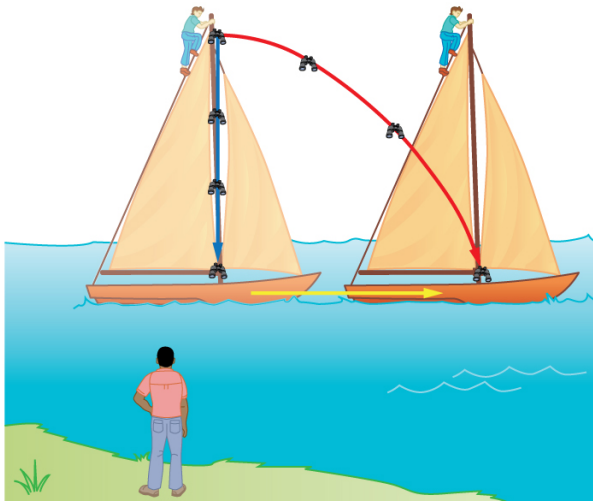
Relative Velocities and Classical Relativity

When adding velocities, we have been careful to specify that the *velocity is relative to some reference frame*. These velocities are called **relative velocities**. For example, the velocity of an airplane relative to an air mass is different from its velocity relative to the ground. Both are quite different from the velocity of an airplane relative to its passengers (which should be close to zero). Relative velocities are one aspect of **relativity**, which is defined to be the study of how different observers moving relative to each other measure the same phenomenon.

Nearly everyone has heard of relativity and immediately associates it with Albert Einstein (1879–1955), the greatest physicist of the 20th century. Einstein revolutionized our view of nature with his *modern* theory of relativity, which we shall study in later chapters. The relative velocities in this section are actually aspects of classical relativity, first discussed correctly by Galileo and Isaac Newton. **Classical relativity** is limited to situations where speeds are less than about 1% of the speed of light—that is, less than 3,000 km/s. Most things we encounter in daily life move slower than this speed.

Let us consider an example of what two different observers see in a situation analyzed long ago by Galileo. Suppose a sailor at the top of a mast on a moving ship drops his binoculars. Where will it hit the deck? Will it hit

at the base of the mast, or will it hit behind the mast because the ship is moving forward? The answer is that if air resistance is negligible, the binoculars will hit at the base of the mast at a point directly below its point of release. Now let us consider what two different observers see when the binoculars drop. One observer is on the ship and the other on shore. The binoculars have no horizontal velocity relative to the observer on the ship, and so he sees them fall straight down the mast. (See [\[link\]](#).) To the observer on shore, the binoculars and the ship have the *same* horizontal velocity, so both move the same distance forward while the binoculars are falling. This observer sees the curved path shown in [\[link\]](#). Although the paths look different to the different observers, each sees the same result—the binoculars hit at the base of the mast and not behind it. To get the correct description, it is crucial to correctly specify the velocities relative to the observer.



Classical relativity. The same motion as viewed by two different observers. An observer on the moving ship sees the binoculars dropped from the top of its mast fall straight down. An observer on shore sees the binoculars take the curved path,

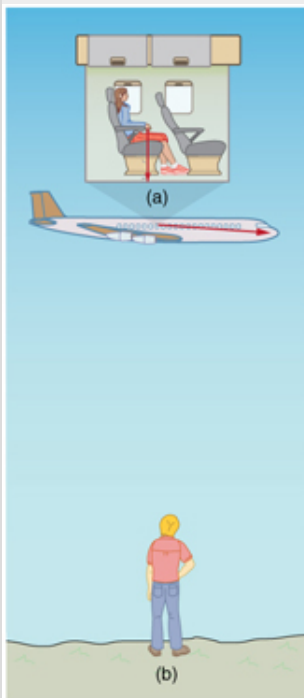
moving forward with the ship.

Both observers see the binoculars strike the deck at the base of the mast. The initial horizontal velocity is different relative to the two observers. (The ship is shown moving rather fast to emphasize the effect.)

Example:

Calculating Relative Velocity: An Airline Passenger Drops a Coin

An airline passenger drops a coin while the plane is moving at 260 m/s. What is the velocity of the coin when it strikes the floor 1.50 m below its point of release: (a) Measured relative to the plane? (b) Measured relative to the Earth?



The motion of

a coin
dropped
inside an
airplane as
viewed by
two different
observers. (a)
An observer
in the plane
sees the coin
fall straight
down. (b) An
observer on
the ground
sees the coin
move almost
horizontally.

Strategy

Both problems can be solved with the techniques for falling objects and projectiles. In part (a), the initial velocity of the coin is zero relative to the plane, so the motion is that of a falling object (one-dimensional). In part (b), the initial velocity is 260 m/s horizontal relative to the Earth and gravity is vertical, so this motion is a projectile motion. In both parts, it is best to use a coordinate system with vertical and horizontal axes.

Solution for (a)

Using the given information, we note that the initial velocity and position are zero, and the final position is 1.50 m. The final velocity can be found using the equation:

Equation:

$$v_y^2 = v_{0y}^2 - 2g(y - y_0).$$

Substituting known values into the equation, we get

Equation:

$$v_y^2 = 0^2 - 2(9.80 \text{ m/s}^2)(-1.50 \text{ m} - 0 \text{ m}) = 29.4 \text{ m}^2/\text{s}^2$$

yielding

Equation:

$$v_y = -5.42 \text{ m/s}.$$

We know that the square root of 29.4 has two roots: 5.42 and -5.42. We choose the negative root because we know that the velocity is directed downwards, and we have defined the positive direction to be upwards. There is no initial horizontal velocity relative to the plane and no horizontal acceleration, and so the motion is straight down relative to the plane.

Solution for (b)

Because the initial vertical velocity is zero relative to the ground and vertical motion is independent of horizontal motion, the final vertical velocity for the coin relative to the ground is $v_y = -5.42 \text{ m/s}$, the same as found in part (a). In contrast to part (a), there now is a horizontal component of the velocity. However, since there is no horizontal acceleration, the initial and final horizontal velocities are the same and $v_x = 260 \text{ m/s}$. The x - and y -components of velocity can be combined to find the magnitude of the final velocity:

Equation:

$$v = \sqrt{v_x^2 + v_y^2}.$$

Thus,

Equation:

$$v = \sqrt{(260 \text{ m/s})^2 + (-5.42 \text{ m/s})^2}$$

yielding

Equation:

$$v = 260.06 \text{ m/s}.$$

The direction is given by:

Equation:

$$\theta = \tan^{-1}(v_y/v_x) = \tan^{-1}(-5.42/260)$$

so that

Equation:

$$\theta = \tan^{-1}(-0.0208) = -1.19^\circ.$$

Discussion

In part (a), the final velocity relative to the plane is the same as it would be if the coin were dropped from rest on the Earth and fell 1.50 m. This result fits our experience; objects in a plane fall the same way when the plane is flying horizontally as when it is at rest on the ground. This result is also true in moving cars. In part (b), an observer on the ground sees a much different motion for the coin. The plane is moving so fast horizontally to begin with that its final velocity is barely greater than the initial velocity. Once again, we see that in two dimensions, vectors do not add like ordinary numbers—the final velocity v in part (b) is *not* $(260 - 5.42)$ m/s; rather, it is 260.06 m/s. The velocity's magnitude had to be calculated to five digits to see any difference from that of the airplane. The motions as seen by different observers (one in the plane and one on the ground) in this example are analogous to those discussed for the binoculars dropped from the mast of a moving ship, except that the velocity of the plane is much larger, so that the two observers see *very* different paths. (See [\[link\]](#).) In addition, both observers see the coin fall 1.50 m vertically, but the one on the ground also sees it move forward 144 m (this calculation is left for the reader). Thus, one observer sees a vertical path, the other a nearly horizontal path.

Note:

Making Connections: Relativity and Einstein

Because Einstein was able to clearly define how measurements are made (some involve light) and because the speed of light is the same for all

observers, the outcomes are spectacularly unexpected. Time varies with observer, energy is stored as increased mass, and more surprises await.

Note:

PhET Explorations: Motion in 2D

Try the new "Motion in 2D" simulation for the latest updated version.

Learn about position, velocity, and acceleration vectors. Move the ball with the mouse or let the simulation move the ball in four types of motion (2 types of linear, simple harmonic, circle).

[Motion in 2D](#)

Summary

- Velocities in two dimensions are added using the same analytical vector techniques, which are rewritten as

Equation:

$$v_x = v \cos \theta$$

Equation:

$$v_y = v \sin \theta$$

Equation:

$$v = \sqrt{v_x^2 + v_y^2}$$

Equation:

$$\theta = \tan^{-1}(v_y/v_x).$$

- Relative velocity is the velocity of an object as observed from a particular reference frame, and it varies dramatically with reference frame.
- **Relativity** is the study of how different observers measure the same phenomenon, particularly when the observers move relative to one another. **Classical relativity** is limited to situations where speed is less than about 1% of the speed of light (3000 km/s).

Conceptual Questions

Exercise:

Problem:

What frame or frames of reference do you instinctively use when driving a car? When flying in a commercial jet airplane?

Exercise:

Problem:

A basketball player dribbling down the court usually keeps his eyes fixed on the players around him. He is moving fast. Why doesn't he need to keep his eyes on the ball?

Exercise:

Problem:

If someone is riding in the back of a pickup truck and throws a softball straight backward, is it possible for the ball to fall straight down as viewed by a person standing at the side of the road? Under what condition would this occur? How would the motion of the ball appear to the person who threw it?

Exercise:**Problem:**

The hat of a jogger running at constant velocity falls off the back of his head. Draw a sketch showing the path of the hat in the jogger's frame of reference. Draw its path as viewed by a stationary observer.

Exercise:**Problem:**

A clod of dirt falls from the bed of a moving truck. It strikes the ground directly below the end of the truck. What is the direction of its velocity relative to the truck just before it hits? Is this the same as the direction of its velocity relative to ground just before it hits? Explain your answers.

Problems & Exercises**Exercise:****Problem:**

Bryan Allen pedaled a human-powered aircraft across the English Channel from the cliffs of Dover to Cap Gris-Nez on June 12, 1979. (a) He flew for 169 min at an average velocity of 3.53 m/s in a direction 45° south of east. What was his total displacement? (b) Allen encountered a headwind averaging 2.00 m/s almost precisely in the opposite direction of his motion relative to the Earth. What was his average velocity relative to the air? (c) What was his total displacement relative to the air mass?

Solution:

(a) 35.8 km, 45° south of east

(b) 5.53 m/s, 45° south of east

(c) 56.1 km, 45° south of east

Exercise:

Problem:

A seagull flies at a velocity of 9.00 m/s straight into the wind. (a) If it takes the bird 20.0 min to travel 6.00 km relative to the Earth, what is the velocity of the wind? (b) If the bird turns around and flies with the wind, how long will he take to return 6.00 km? (c) Discuss how the wind affects the total round-trip time compared to what it would be with no wind.

Exercise:

Problem:

Near the end of a marathon race, the first two runners are separated by a distance of 45.0 m. The front runner has a velocity of 3.50 m/s, and the second a velocity of 4.20 m/s. (a) What is the velocity of the second runner relative to the first? (b) If the front runner is 250 m from the finish line, who will win the race, assuming they run at constant velocity? (c) What distance ahead will the winner be when she crosses the finish line?

Solution:

(a) 0.70 m/s faster

(b) Second runner wins

(c) 4.17 m

Exercise:

Problem:

Verify that the coin dropped by the airline passenger in the [\[link\]](#) travels 144 m horizontally while falling 1.50 m in the frame of reference of the Earth.

Exercise:

Problem:

A football quarterback is moving straight backward at a speed of 2.00 m/s when he throws a pass to a player 18.0 m straight downfield. The ball is thrown at an angle of 25.0° relative to the ground and is caught at the same height as it is released. What is the initial velocity of the ball *relative to the quarterback* ?

Solution:

17.0 m/s, 22.1°

Exercise:**Problem:**

A ship sets sail from Rotterdam, The Netherlands, heading due north at 7.00 m/s relative to the water. The local ocean current is 1.50 m/s in a direction 40.0° north of east. What is the velocity of the ship relative to the Earth?

Exercise:**Problem:**

(a) A jet airplane flying from Darwin, Australia, has an air speed of 260 m/s in a direction 5.0° south of west. It is in the jet stream, which is blowing at 35.0 m/s in a direction 15° south of east. What is the velocity of the airplane relative to the Earth? (b) Discuss whether your answers are consistent with your expectations for the effect of the wind on the plane's path.

Solution:

(a) 230 m/s, 8.0° south of west

(b) The wind should make the plane travel slower and more to the south, which is what was calculated.

Exercise:

Problem:

(a) In what direction would the ship in [\[link\]](#) have to travel in order to have a velocity straight north relative to the Earth, assuming its speed relative to the water remains 7.00 m/s ? (b) What would its speed be relative to the Earth?

Exercise:**Problem:**

(a) Another airplane is flying in a jet stream that is blowing at 45.0 m/s in a direction 20° south of east (as in [\[link\]](#)). Its direction of motion relative to the Earth is 45.0° south of west, while its direction of travel relative to the air is 5.00° south of west. What is the airplane's speed relative to the air mass? (b) What is the airplane's speed relative to the Earth?

Solution:

(a) 63.5 m/s

(b) 29.6 m/s

Exercise:**Problem:**

A sandal is dropped from the top of a 15.0-m -high mast on a ship moving at 1.75 m/s due south. Calculate the velocity of the sandal when it hits the deck of the ship: (a) relative to the ship and (b) relative to a stationary observer on shore. (c) Discuss how the answers give a consistent result for the position at which the sandal hits the deck.

Exercise:

Problem:

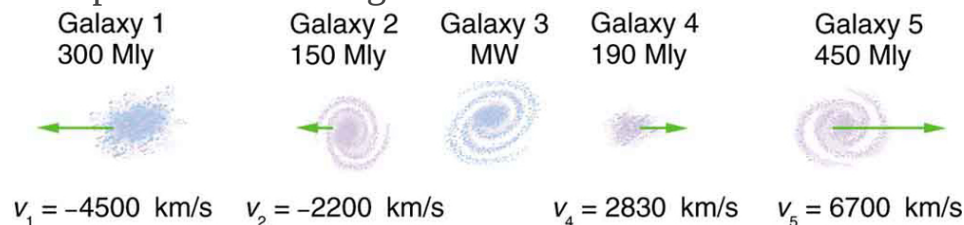
The velocity of the wind relative to the water is crucial to sailboats. Suppose a sailboat is in an ocean current that has a velocity of 2.20 m/s in a direction 30.0° east of north relative to the Earth. It encounters a wind that has a velocity of 4.50 m/s in a direction of 50.0° south of west relative to the Earth. What is the velocity of the wind relative to the water?

Solution:

6.68 m/s, 53.3° south of west

Exercise:**Problem:**

The great astronomer Edwin Hubble discovered that all distant galaxies are receding from our Milky Way Galaxy with velocities proportional to their distances. [\[link\]](#) illustrates this for five galaxies lying along a straight line, with the Milky Way Galaxy at the center. Using the data from the figure, calculate the velocities: (a) relative to galaxy 2 and (b) relative to galaxy 5. The results mean that observers on all galaxies will see themselves at the center of the expanding universe, and they would likely be aware of relative velocities, concluding that it is not possible to locate the center of expansion with the given information.



Five galaxies on a straight line, showing their distances and velocities relative to the Milky Way (MW) Galaxy. The distances are in millions of light years (Mly), where a light year is the distance light

travels in one year. The velocities are nearly proportional to the distances. The sizes of the galaxies are greatly exaggerated; an average galaxy is about 0.1 Mly across.

Exercise:

Problem:

- (a) Use the distance and velocity data in [\[link\]](#) to find the rate of expansion as a function of distance.
- (b) If you extrapolate back in time, how long ago would all of the galaxies have been at approximately the same position? The two parts of this problem give you some idea of how the Hubble constant for universal expansion and the time back to the Big Bang are determined, respectively.

Solution:

(a) $H_{\text{average}} = 14.9 \frac{\text{km/s}}{\text{Mly}}$

(b) 20.2 billion years

Exercise:

Problem:

An athlete crosses a 25-m-wide river by swimming perpendicular to the water current at a speed of 0.5 m/s relative to the water. He reaches the opposite side at a distance 40 m downstream from his starting point. How fast is the water in the river flowing with respect to the ground? What is the speed of the swimmer with respect to a friend at rest on the ground?

Exercise:

Problem:

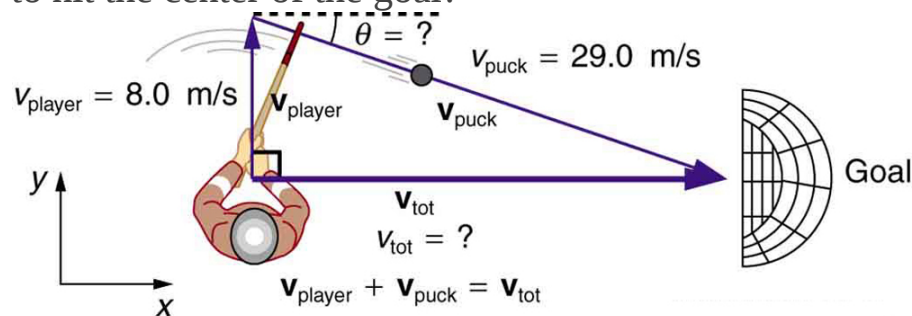
A ship sailing in the Gulf Stream is heading 25.0° west of north at a speed of 4.00 m/s relative to the water. Its velocity relative to the Earth is 4.80 m/s 5.00° west of north. What is the velocity of the Gulf Stream? (The velocity obtained is typical for the Gulf Stream a few hundred kilometers off the east coast of the United States.)

Solution:

1.72 m/s , 42.3° north of east

Exercise:**Problem:**

An ice hockey player is moving at 8.00 m/s when he hits the puck toward the goal. The speed of the puck relative to the player is 29.0 m/s . The line between the center of the goal and the player makes a 90.0° angle relative to his path as shown in [\[link\]](#). What angle must the puck's velocity make relative to the player (in his frame of reference) to hit the center of the goal?



An ice hockey player moving across the rink must shoot backward to give the puck a velocity toward the goal.

Exercise:

Problem:

Unreasonable Results Suppose you wish to shoot supplies straight up to astronauts in an orbit 36,000 km above the surface of the Earth. (a) At what velocity must the supplies be launched? (b) What is unreasonable about this velocity? (c) Is there a problem with the relative velocity between the supplies and the astronauts when the supplies reach their maximum height? (d) Is the premise unreasonable or is the available equation inapplicable? Explain your answer.

Exercise:**Problem:**

Unreasonable Results A commercial airplane has an air speed of 280 m/s due east and flies with a strong tailwind. It travels 3000 km in a direction 5° south of east in 1.50 h. (a) What was the velocity of the plane relative to the ground? (b) Calculate the magnitude and direction of the tailwind's velocity. (c) What is unreasonable about both of these velocities? (d) Which premise is unreasonable?

Exercise:**Problem:**

Construct Your Own Problem Consider an airplane headed for a runway in a cross wind. Construct a problem in which you calculate the angle the airplane must fly relative to the air mass in order to have a velocity parallel to the runway. Among the things to consider are the direction of the runway, the wind speed and direction (its velocity) and the speed of the plane relative to the air mass. Also calculate the speed of the airplane relative to the ground. Discuss any last minute maneuvers the pilot might have to perform in order for the plane to land with its wheels pointing straight down the runway.

Glossary

classical relativity

the study of relative velocities in situations where speeds are less than about 1% of the speed of light—that is, less than 3000 km/s

relative velocity

the velocity of an object as observed from a particular reference frame

relativity

the study of how different observers moving relative to each other measure the same phenomenon

velocity

speed in a given direction

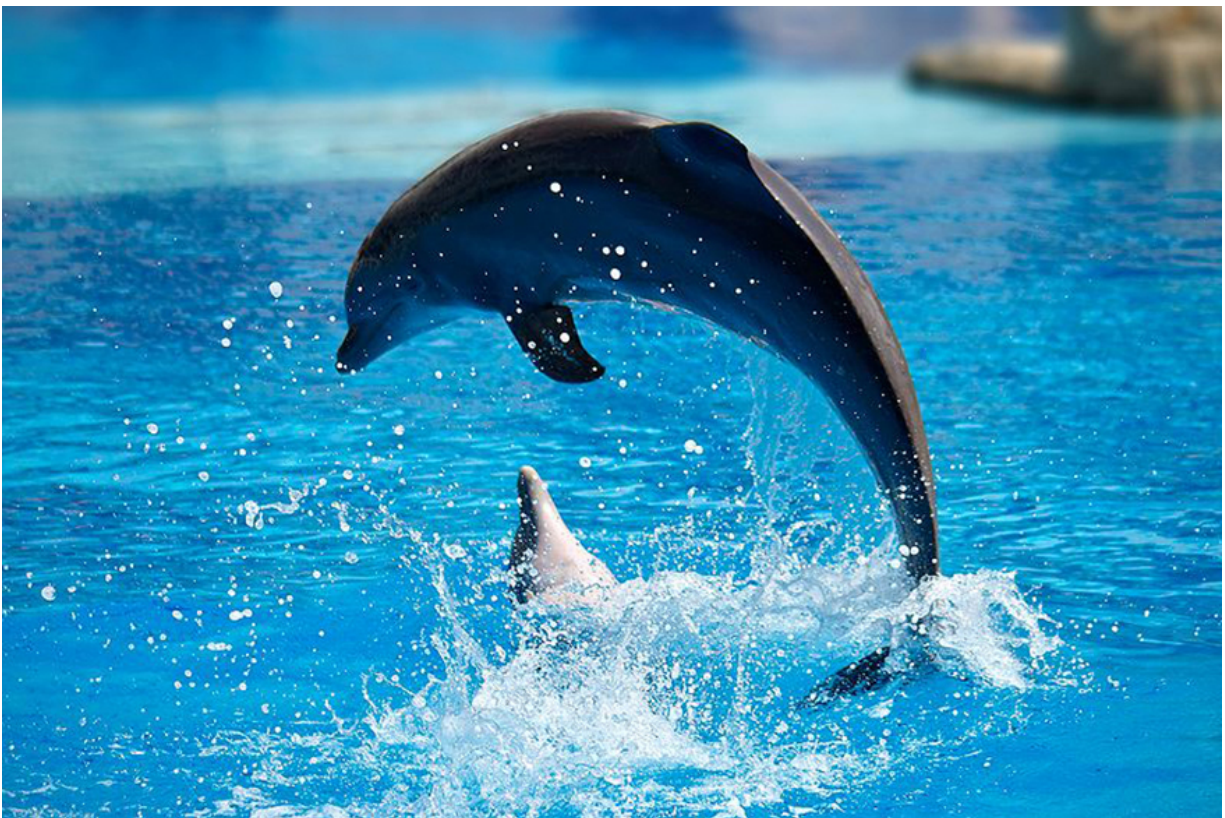
vector addition

the rules that apply to adding vectors together

Introduction to Dynamics: Newton's Laws of Motion

class="introduction"

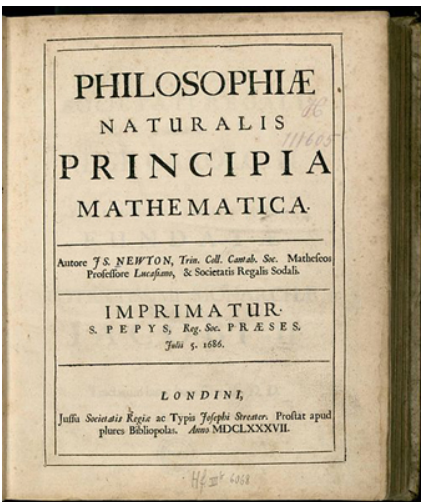
Newton's laws of motion describe the motion of the dolphin's path.
(credit: Jin Jang)



Motion draws our attention. Motion itself can be beautiful, causing us to marvel at the forces needed to achieve spectacular motion, such as that of a

dolphin jumping out of the water, or a pole vaulter, or the flight of a bird, or the orbit of a satellite. The study of motion is kinematics, but kinematics only *describes* the way objects move—their velocity and their acceleration. **Dynamics** considers the forces that affect the motion of moving objects and systems. Newton's laws of motion are the foundation of dynamics. These laws provide an example of the breadth and simplicity of principles under which nature functions. They are also universal laws in that they apply to similar situations on Earth as well as in space.

Isaac Newton's (1642–1727) laws of motion were just one part of the monumental work that has made him legendary. The development of Newton's laws marks the transition from the Renaissance into the modern era. This transition was characterized by a revolutionary change in the way people thought about the physical universe. For many centuries natural philosophers had debated the nature of the universe based largely on certain rules of logic with great weight given to the thoughts of earlier classical philosophers such as Aristotle (384–322 BC). Among the many great thinkers who contributed to this change were Newton and Galileo.



Isaac Newton's
monumental work,
*Philosophiæ
Naturalis Principia
Mathematica*, was
published in 1687. It
proposed scientific

laws that are still
used today to
describe the motion
of objects. (credit:
Service commun de
la documentation de
l'Université de
Strasbourg)

Galileo was instrumental in establishing *observation* as the absolute determinant of truth, rather than “logical” argument. Galileo’s use of the telescope was his most notable achievement in demonstrating the importance of observation. He discovered moons orbiting Jupiter and made other observations that were inconsistent with certain ancient ideas and religious dogma. For this reason, and because of the manner in which he dealt with those in authority, Galileo was tried by the Inquisition and punished. He spent the final years of his life under a form of house arrest. Because others before Galileo had also made discoveries by *observing* the nature of the universe, and because repeated observations verified those of Galileo, his work could not be suppressed or denied. After his death, his work was verified by others, and his ideas were eventually accepted by the church and scientific communities.

Galileo also contributed to the formation of what is now called Newton’s first law of motion. Newton made use of the work of his predecessors, which enabled him to develop laws of motion, discover the law of gravity, invent calculus, and make great contributions to the theories of light and color. It is amazing that many of these developments were made with Newton working alone, without the benefit of the usual interactions that take place among scientists today.

It was not until the advent of modern physics early in the 20th century that it was discovered that Newton’s laws of motion produce a good approximation to motion only when the objects are moving at speeds much, much less than the speed of light and when those objects are larger than the

size of most molecules (about 10^{-9} m in diameter). These constraints define the realm of classical mechanics, as discussed in [Introduction to the Nature of Science and Physics](#). At the beginning of the 20th century, Albert Einstein (1879–1955) developed the theory of relativity and, along with many other scientists, developed quantum theory. This theory does not have the constraints present in classical physics. All of the situations we consider in this chapter, and all those preceding the introduction of relativity in [Special Relativity](#), are in the realm of classical physics.

Note:

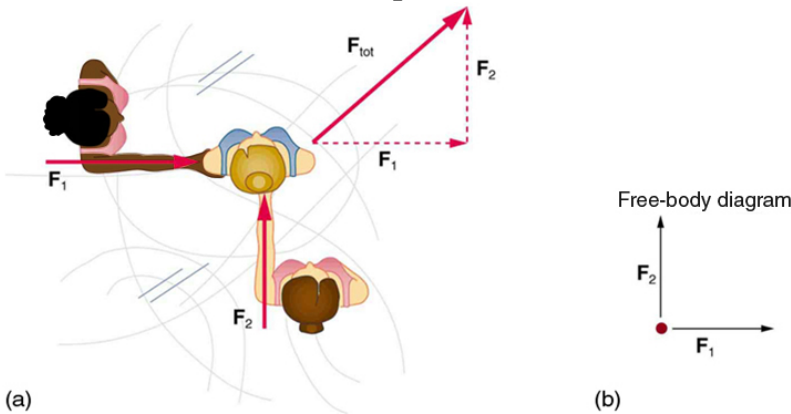
Making Connections: Past and Present Philosophy

The importance of observation and the concept of *cause and effect* were not always so entrenched in human thinking. This realization was a part of the evolution of modern physics from natural philosophy. The achievements of Galileo, Newton, Einstein, and others were key milestones in the history of scientific thought. Most of the scientific theories that are described in this book descended from the work of these scientists.

Development of Force Concept

- Understand the definition of force.

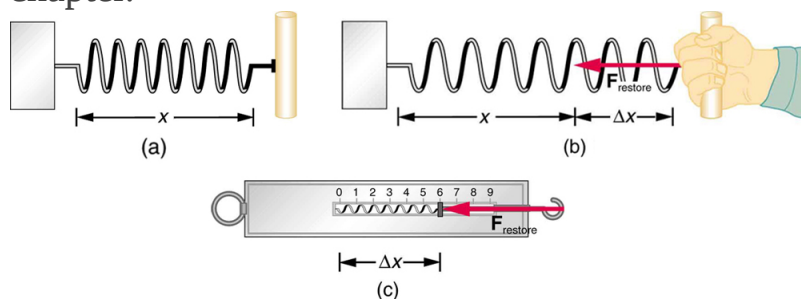
Dynamics is the study of the forces that cause objects and systems to move. To understand this, we need a working definition of force. Our intuitive definition of **force**—that is, a push or a pull—is a good place to start. We know that a push or pull has both magnitude and direction (therefore, it is a vector quantity) and can vary considerably in each regard. For example, a cannon exerts a strong force on a cannonball that is launched into the air. In contrast, Earth exerts only a tiny downward pull on a flea. Our everyday experiences also give us a good idea of how multiple forces add. If two people push in different directions on a third person, as illustrated in [\[link\]](#), we might expect the total force to be in the direction shown. Since force is a vector, it adds just like other vectors, as illustrated in [\[link\]](#)(a) for two ice skaters. Forces, like other vectors, are represented by arrows and can be added using the familiar head-to-tail method or by trigonometric methods. These ideas were developed in [Two-Dimensional Kinematics](#).



Part (a) shows an overhead view of two ice skaters pushing on a third. Forces are vectors and add like other vectors, so the total force on the third skater is in the direction shown. In part (b), we see a free-body diagram representing the forces acting on the third skater.

[\[link\]](#)(b) is our first example of a **free-body diagram**, which is a technique used to illustrate all the **external forces** acting on a body. The body is represented by a single isolated point (or free body), and only those forces acting *on* the body from the outside (external forces) are shown. (These forces are the only ones shown, because only external forces acting on the body affect its motion. We can ignore any internal forces within the body.) Free-body diagrams are very useful in analyzing forces acting on a system and are employed extensively in the study and application of Newton's laws of motion.

A more quantitative definition of force can be based on some standard force, just as distance is measured in units relative to a standard distance. One possibility is to stretch a spring a certain fixed distance, as illustrated in [\[link\]](#), and use the force it exerts to pull itself back to its relaxed shape—called a *restoring force*—as a standard. The magnitude of all other forces can be stated as multiples of this standard unit of force. Many other possibilities exist for standard forces. (One that we will encounter in [Magnetism](#) is the magnetic force between two wires carrying electric current.) Some alternative definitions of force will be given later in this chapter.



The force exerted by a stretched spring can be used as a standard unit of force. (a) This spring has a length x when undistorted. (b) When stretched a distance Δx , the spring exerts a restoring force, $\mathbf{F}_{\text{restore}}$, which is reproducible. (c) A spring scale is one device that uses a spring to measure force. The force $\mathbf{F}_{\text{restore}}$ is exerted on whatever is attached to the hook. Here $\mathbf{F}_{\text{restore}}$ has a

magnitude of 6 units in the force standard being employed.

Note:

Take-Home Experiment: Force Standards

To investigate force standards and cause and effect, get two identical rubber bands. Hang one rubber band vertically on a hook. Find a small household item that could be attached to the rubber band using a paper clip, and use this item as a weight to investigate the stretch of the rubber band. Measure the amount of stretch produced in the rubber band with one, two, and four of these (identical) items suspended from the rubber band. What is the relationship between the number of items and the amount of stretch? How large a stretch would you expect for the same number of items suspended from two rubber bands? What happens to the amount of stretch of the rubber band (with the weights attached) if the weights are also pushed to the side with a pencil?

Section Summary

- **Dynamics** is the study of how forces affect the motion of objects.
- **Force** is a push or pull that can be defined in terms of various standards, and it is a vector having both magnitude and direction.
- **External forces** are any outside forces that act on a body. A **free-body diagram** is a drawing of all external forces acting on a body.

Conceptual Questions

Exercise:

Problem:

Propose a force standard different from the example of a stretched spring discussed in the text. Your standard must be capable of producing the same force repeatedly.

Exercise:**Problem:**

What properties do forces have that allow us to classify them as vectors?

Glossary

dynamics

the study of how forces affect the motion of objects and systems

external force

a force acting on an object or system that originates outside of the object or system

free-body diagram

a sketch showing all of the external forces acting on an object or system; the system is represented by a dot, and the forces are represented by vectors extending outward from the dot

force

a push or pull on an object with a specific magnitude and direction; can be represented by vectors; can be expressed as a multiple of a standard force

Newton's First Law of Motion: Inertia

- Define mass and inertia.
- Understand Newton's first law of motion.

Experience suggests that an object at rest will remain at rest if left alone, and that an object in motion tends to slow down and stop unless some effort is made to keep it moving. What **Newton's first law of motion** states, however, is the following:

Note:

Newton's First Law of Motion

A body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force.

Note the repeated use of the verb “remains.” We can think of this law as preserving the status quo of motion.

Rather than contradicting our experience, **Newton's first law of motion** states that there must be a *cause* (which is a net external force) *for there to be any change in velocity (either a change in magnitude or direction)*. We will define *net external force* in the next section. An object sliding across a table or floor slows down due to the net force of friction acting on the object. If friction disappeared, would the object still slow down?

The idea of cause and effect is crucial in accurately describing what happens in various situations. For example, consider what happens to an object sliding along a rough horizontal surface. The object quickly grinds to a halt. If we spray the surface with talcum powder to make the surface smoother, the object slides farther. If we make the surface even smoother by rubbing lubricating oil on it, the object slides farther yet. Extrapolating to a frictionless surface, we can imagine the object sliding in a straight line indefinitely. Friction is thus the *cause* of the slowing (consistent with Newton's first law). The object would not slow down at all if friction were

completely eliminated. Consider an air hockey table. When the air is turned off, the puck slides only a short distance before friction slows it to a stop. However, when the air is turned on, it creates a nearly frictionless surface, and the puck glides long distances without slowing down. Additionally, if we know enough about the friction, we can accurately predict how quickly the object will slow down. Friction is an external force.

Newton's first law is completely general and can be applied to anything from an object sliding on a table to a satellite in orbit to blood pumped from the heart. Experiments have thoroughly verified that any change in velocity (speed or direction) must be caused by an external force. The idea of *generally applicable or universal laws* is important not only here—it is a basic feature of all laws of physics. Identifying these laws is like recognizing patterns in nature from which further patterns can be discovered. The genius of Galileo, who first developed the idea for the first law, and Newton, who clarified it, was to ask the fundamental question, “What is the cause?” Thinking in terms of cause and effect is a worldview fundamentally different from the typical ancient Greek approach when questions such as “Why does a tiger have stripes?” would have been answered in Aristotelian fashion, “That is the nature of the beast.” True perhaps, but not a useful insight.

Mass

The property of a body to remain at rest or to remain in motion with constant velocity is called **inertia**. Newton's first law is often called the **law of inertia**. As we know from experience, some objects have more inertia than others. It is obviously more difficult to change the motion of a large boulder than that of a basketball, for example. The inertia of an object is measured by its **mass**. Roughly speaking, mass is a measure of the amount of “stuff” (or matter) in something. The quantity or amount of matter in an object is determined by the numbers of atoms and molecules of various types it contains. Unlike weight, mass does not vary with location. The mass of an object is the same on Earth, in orbit, or on the surface of the Moon. In practice, it is very difficult to count and identify all of the atoms and molecules in an object, so masses are not often determined in this

manner. Operationally, the masses of objects are determined by comparison with the standard kilogram.

Exercise:

Check Your Understanding

Problem:

Which has more mass: a kilogram of cotton balls or a kilogram of gold?

Solution:

Answer

They are equal. A kilogram of one substance is equal in mass to a kilogram of another substance. The quantities that might differ between them are volume and density.

Section Summary

- **Newton's first law of motion** states that a body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force. This is also known as the **law of inertia**.
- **Inertia** is the tendency of an object to remain at rest or remain in motion. Inertia is related to an object's mass.
- **Mass** is the quantity of matter in a substance.

Conceptual Questions

Exercise:

Problem: How are inertia and mass related?

Exercise:

Problem:

What is the relationship between weight and mass? Which is an intrinsic, unchanging property of a body?

Glossary

inertia

the tendency of an object to remain at rest or remain in motion

law of inertia

see Newton's first law of motion

mass

the quantity of matter in a substance; measured in kilograms

Newton's first law of motion

a body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force; also known as the law of inertia

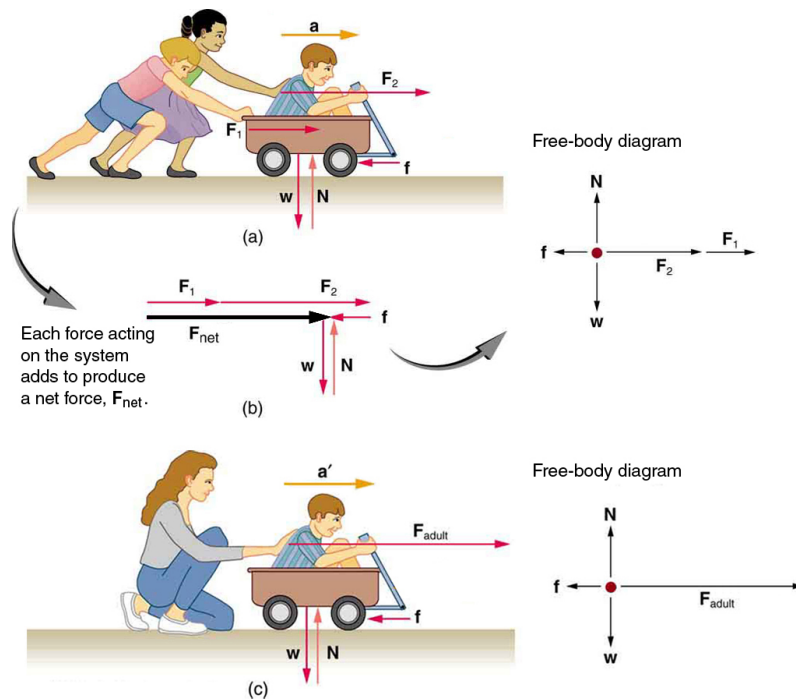
Newton's Second Law of Motion: Concept of a System

- Define net force, external force, and system.
- Understand Newton's second law of motion.
- Apply Newton's second law to determine the weight of an object.

Newton's second law of motion is closely related to Newton's first law of motion. It mathematically states the cause and effect relationship between force and changes in motion. Newton's second law of motion is more quantitative and is used extensively to calculate what happens in situations involving a force. Before we can write down Newton's second law as a simple equation giving the exact relationship of force, mass, and acceleration, we need to sharpen some ideas that have already been mentioned.

First, what do we mean by a change in motion? The answer is that a change in motion is equivalent to a change in velocity. A change in velocity means, by definition, that there is an **acceleration**. Newton's first law says that a net external force causes a change in motion; thus, we see that a *net external force causes acceleration*.

Another question immediately arises. What do we mean by an external force? An intuitive notion of external is correct—an **external force** acts from outside the **system** of interest. For example, in [\[link\]](#)(a) the system of interest is the wagon plus the child in it. The two forces exerted by the other children are external forces. An internal force acts between elements of the system. Again looking at [\[link\]](#)(a), the force the child in the wagon exerts to hang onto the wagon is an internal force between elements of the system of interest. Only external forces affect the motion of a system, according to Newton's first law. (The internal forces actually cancel, as we shall see in the next section.) *You must define the boundaries of the system before you can determine which forces are external.* Sometimes the system is obvious, whereas other times identifying the boundaries of a system is more subtle. The concept of a system is fundamental to many areas of physics, as is the correct application of Newton's laws. This concept will be revisited many times on our journey through physics.



Different forces exerted on the same mass produce different accelerations. (a) Two children push a wagon with a child in it. Arrows representing all external forces are shown. The system of interest is the wagon and its rider. The weight w of the system and the support of the ground N are also shown for completeness and are assumed to cancel. The vector f represents the friction acting on the wagon, and it acts to the left, opposing the motion of the wagon. (b) All of the external forces acting on the system add together to produce a net force, F_{net} . The free-body diagram shows all of the forces acting on the system of interest. The dot represents the center of mass of the system. Each force vector extends from this dot. Because there are two forces acting to the right, we draw the vectors collinearly. (c) A larger net external force produces a larger

acceleration ($\mathbf{a}' > \mathbf{a}$) when an adult pushes the child.

Now, it seems reasonable that acceleration should be directly proportional to and in the same direction as the net (total) external force acting on a system. This assumption has been verified experimentally and is illustrated in [\[link\]](#). In part (a), a smaller force causes a smaller acceleration than the larger force illustrated in part (c). For completeness, the vertical forces are also shown; they are assumed to cancel since there is no acceleration in the vertical direction. The vertical forces are the weight \mathbf{w} and the support of the ground \mathbf{N} , and the horizontal force \mathbf{f} represents the force of friction. These will be discussed in more detail in later sections. For now, we will define **friction** as a force that opposes the motion past each other of objects that are touching. [\[link\]](#)(b) shows how vectors representing the external forces add together to produce a net force, \mathbf{F}_{net} .

To obtain an equation for Newton's second law, we first write the relationship of acceleration and net external force as the proportionality **Equation:**

$$\mathbf{a} \propto \mathbf{F}_{\text{net}},$$

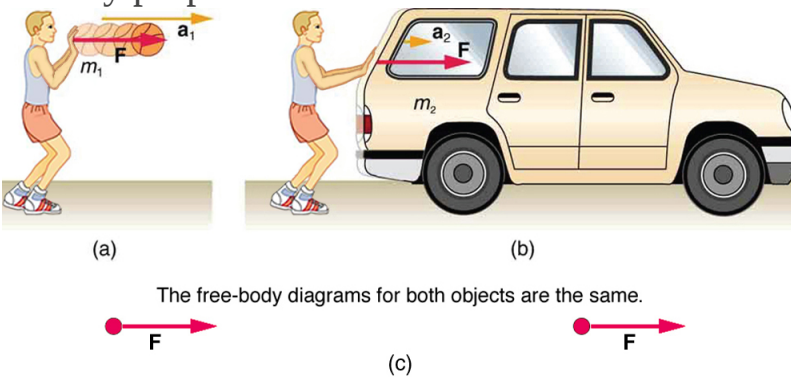
where the symbol \propto means “proportional to,” and \mathbf{F}_{net} is the **net external force**. (The net external force is the vector sum of all external forces and can be determined graphically, using the head-to-tail method, or analytically, using components. The techniques are the same as for the addition of other vectors, and are covered in [Two-Dimensional Kinematics](#).) This proportionality states what we have said in words—*acceleration is directly proportional to the net external force*. Once the system of interest is chosen, it is important to identify the external forces and ignore the internal ones. It is a tremendous simplification not to have to consider the numerous internal forces acting between objects within the system, such as muscular forces within the child's body, let alone the myriad of forces between atoms in the objects, but by doing so, we can easily solve some very complex problems with only minimal error due to our simplification

Now, it also seems reasonable that acceleration should be inversely proportional to the mass of the system. In other words, the larger the mass (the inertia), the smaller the acceleration produced by a given force. And indeed, as illustrated in [\[link\]](#), the same net external force applied to a car produces a much smaller acceleration than when applied to a basketball. The proportionality is written as

Equation:

$$\mathbf{a} \propto \frac{1}{m}$$

where m is the mass of the system. Experiments have shown that acceleration is exactly inversely proportional to mass, just as it is exactly linearly proportional to the net external force.



The same force exerted on systems of different masses produces different accelerations. (a) A basketball player pushes on a basketball to make a pass. (The effect of gravity on the ball is ignored.) (b) The same player exerts an identical force on a stalled SUV and produces a far smaller acceleration (even if friction is negligible). (c) The free-body diagrams are identical, permitting direct comparison of the two situations. A series of patterns for the free-body diagram will emerge as you do more problems.

It has been found that the acceleration of an object depends *only* on the net external force and the mass of the object. Combining the two proportionalities just given yields Newton's second law of motion.

Note:

Newton's Second Law of Motion

The acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system, and inversely proportional to its mass.

In equation form, Newton's second law of motion is

Equation:

$$\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m}.$$

This is often written in the more familiar form

Equation:

$$\mathbf{F}_{\text{net}} = m\mathbf{a}.$$

When only the magnitude of force and acceleration are considered, this equation is simply

Equation:

$$F_{\text{net}} = ma.$$

Although these last two equations are really the same, the first gives more insight into what Newton's second law means. The law is a *cause and effect relationship* among three quantities that is not simply based on their definitions. The validity of the second law is completely based on experimental verification.

Units of Force

$\mathbf{F}_{\text{net}} = m\mathbf{a}$ is used to define the units of force in terms of the three basic units for mass, length, and time. The SI unit of force is called the **newton** (abbreviated N) and is the force needed to accelerate a 1-kg system at the rate of 1m/s^2 . That is, since $\mathbf{F}_{\text{net}} = m\mathbf{a}$,

Equation:

$$1 \text{ N} = 1 \text{ kg} \cdot \text{m/s}^2.$$

While almost the entire world uses the newton for the unit of force, in the United States the most familiar unit of force is the pound (lb), where $1 \text{ N} = 0.225 \text{ lb}$.

Weight and the Gravitational Force

When an object is dropped, it accelerates toward the center of Earth. Newton's second law states that a net force on an object is responsible for its acceleration. If air resistance is negligible, the net force on a falling object is the gravitational force, commonly called its **weight \mathbf{w}** . Weight can be denoted as a vector \mathbf{w} because it has a direction; *down* is, by definition, the direction of gravity, and hence weight is a downward force. The magnitude of weight is denoted as w . Galileo was instrumental in showing that, in the absence of air resistance, all objects fall with the same acceleration g . Using Galileo's result and Newton's second law, we can derive an equation for weight.

Consider an object with mass m falling downward toward Earth. It experiences only the downward force of gravity, which has magnitude w . Newton's second law states that the magnitude of the net external force on an object is $F_{\text{net}} = ma$.

Since the object experiences only the downward force of gravity, $F_{\text{net}} = w$. We know that the acceleration of an object due to gravity is g , or $a = g$. Substituting these into Newton's second law gives

Note:**Weight**

This is the equation for *weight*—the gravitational force on a mass m :

Equation:

$$w = mg.$$

Since $g = 9.80 \text{ m/s}^2$ on Earth, the weight of a 1.0 kg object on Earth is 9.8 N, as we see:

Equation:

$$w = mg = (1.0 \text{ kg})(9.80 \text{ m/s}^2) = 9.8 \text{ N}.$$

Recall that g can take a positive or negative value, depending on the positive direction in the coordinate system. Be sure to take this into consideration when solving problems with weight.

When the net external force on an object is its weight, we say that it is in **free-fall**. That is, the only force acting on the object is the force of gravity. In the real world, when objects fall downward toward Earth, they are never truly in free-fall because there is always some upward force from the air acting on the object.

The acceleration due to gravity g varies slightly over the surface of Earth, so that the weight of an object depends on location and is not an intrinsic property of the object. Weight varies dramatically if one leaves Earth's surface. On the Moon, for example, the acceleration due to gravity is only 1.67 m/s^2 . A 1.0-kg mass thus has a weight of 9.8 N on Earth and only about 1.7 N on the Moon.

The broadest definition of weight in this sense is that *the weight of an object is the gravitational force on it from the nearest large body*, such as Earth, the Moon, the Sun, and so on. This is the most common and useful definition of weight in physics. It differs dramatically, however, from the definition of weight used by NASA and the popular media in relation to space travel and exploration. When they speak of “weightlessness” and

“microgravity,” they are really referring to the phenomenon we call “free-fall” in physics. We shall use the above definition of weight, and we will make careful distinctions between free-fall and actual weightlessness.

It is important to be aware that weight and mass are very different physical quantities, although they are closely related. Mass is the quantity of matter (how much “stuff”) and does not vary in classical physics, whereas weight is the gravitational force and does vary depending on gravity. It is tempting to equate the two, since most of our examples take place on Earth, where the weight of an object only varies a little with the location of the object. Furthermore, the terms *mass* and *weight* are used interchangeably in everyday language; for example, our medical records often show our “weight” in kilograms, but never in the correct units of newtons.

Note:**Common Misconceptions: Mass vs. Weight**

Mass and weight are often used interchangeably in everyday language. However, in science, these terms are distinctly different from one another. Mass is a measure of how much matter is in an object. The typical measure of mass is the kilogram (or the “slug” in English units). Weight, on the other hand, is a measure of the force of gravity acting on an object. Weight is equal to the mass of an object (m) multiplied by the acceleration due to gravity (g). Like any other force, weight is measured in terms of newtons (or pounds in English units).

Assuming the mass of an object is kept intact, it will remain the same, regardless of its location. However, because weight depends on the acceleration due to gravity, the weight of an object *can change* when the object enters into a region with stronger or weaker gravity. For example, the acceleration due to gravity on the Moon is 1.67 m/s^2 (which is much less than the acceleration due to gravity on Earth, 9.80 m/s^2). If you measured your weight on Earth and then measured your weight on the Moon, you would find that you “weigh” much less, even though you do not look any skinnier. This is because the force of gravity is weaker on the Moon. In fact, when people say that they are “losing weight,” they really

mean that they are losing “mass” (which in turn causes them to weigh less).

Note:

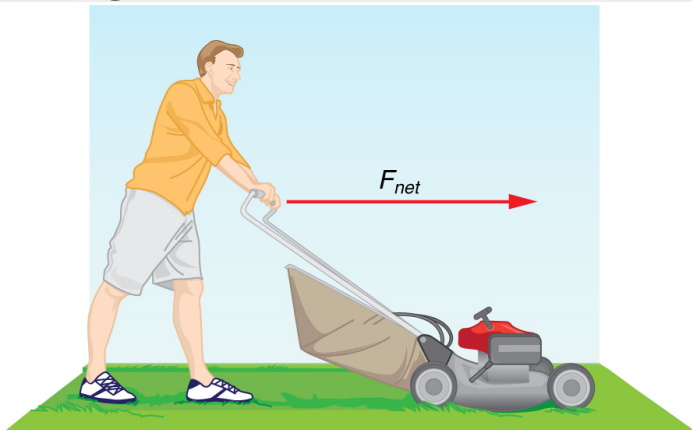
Take-Home Experiment: Mass and Weight

What do bathroom scales measure? When you stand on a bathroom scale, what happens to the scale? It depresses slightly. The scale contains springs that compress in proportion to your weight—similar to rubber bands expanding when pulled. The springs provide a measure of your weight (for an object which is not accelerating). This is a force in newtons (or pounds). In most countries, the measurement is divided by 9.80 to give a reading in mass units of kilograms. The scale measures weight but is calibrated to provide information about mass. While standing on a bathroom scale, push down on a table next to you. What happens to the reading? Why? Would your scale measure the same “mass” on Earth as on the Moon?

Example:

What Acceleration Can a Person Produce when Pushing a Lawn Mower?

Suppose that the net external force (push minus friction) exerted on a lawn mower is 51 N (about 11 lb) parallel to the ground. The mass of the mower is 24 kg. What is its acceleration?



The net force on a lawn mower is 51

N to the right. At what rate does the lawn mower accelerate to the right?

Strategy

Since \mathbf{F}_{net} and m are given, the acceleration can be calculated directly from Newton's second law as stated in $\mathbf{F}_{\text{net}} = m\mathbf{a}$.

Solution

The magnitude of the acceleration a is $a = \frac{F_{\text{net}}}{m}$. Entering known values gives

Equation:

$$a = \frac{51 \text{ N}}{24 \text{ kg}}$$

Substituting the units $\text{kg} \cdot \text{m}/\text{s}^2$ for N yields

Equation:

$$a = \frac{51 \text{ kg} \cdot \text{m}/\text{s}^2}{24 \text{ kg}} = 2.1 \text{ m}/\text{s}^2.$$

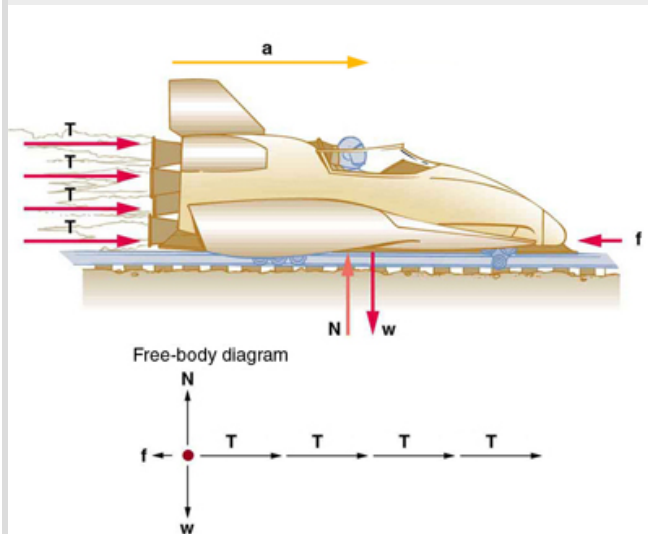
Discussion

The direction of the acceleration is the same direction as that of the net force, which is parallel to the ground. There is no information given in this example about the individual external forces acting on the system, but we can say something about their relative magnitudes. For example, the force exerted by the person pushing the mower must be greater than the friction opposing the motion (since we know the mower moves forward), and the vertical forces must cancel if there is to be no acceleration in the vertical direction (the mower is moving only horizontally). The acceleration found is small enough to be reasonable for a person pushing a mower. Such an effort would not last too long because the person's top speed would soon be reached.

Example:

What Rocket Thrust Accelerates This Sled?

Prior to manned space flights, rocket sleds were used to test aircraft, missile equipment, and physiological effects on human subjects at high speeds. They consisted of a platform that was mounted on one or two rails and propelled by several rockets. Calculate the magnitude of force exerted by each rocket, called its thrust \mathbf{T} , for the four-rocket propulsion system shown in [\[link\]](#). The sled's initial acceleration is 49 m/s^2 , the mass of the system is 2100 kg , and the force of friction opposing the motion is known to be 650 N .



A sled experiences a rocket thrust that accelerates it to the right. Each rocket creates an identical thrust \mathbf{T} . As in other situations where there is only horizontal acceleration, the vertical forces cancel. The ground exerts an upward force \mathbf{N} on the system that is equal in magnitude and opposite in direction to its weight, \mathbf{w} . The system here is the sled, its rockets, and rider, so none of the forces *between* these objects are considered. The arrow representing friction (\mathbf{f}) is drawn larger than scale.

Strategy

Although there are forces acting vertically and horizontally, we assume the vertical forces cancel since there is no vertical acceleration. This leaves us with only horizontal forces and a simpler one-dimensional problem.

Directions are indicated with plus or minus signs, with right taken as the positive direction. See the free-body diagram in the figure.

Solution

Since acceleration, mass, and the force of friction are given, we start with Newton's second law and look for ways to find the thrust of the engines.

Since we have defined the direction of the force and acceleration as acting "to the right," we need to consider only the magnitudes of these quantities in the calculations. Hence we begin with

Equation:

$$F_{\text{net}} = ma,$$

where F_{net} is the net force along the horizontal direction. We can see from [\[link\]](#) that the engine thrusts add, while friction opposes the thrust. In equation form, the net external force is

Equation:

$$F_{\text{net}} = 4T - f.$$

Substituting this into Newton's second law gives

Equation:

$$F_{\text{net}} = ma = 4T - f.$$

Using a little algebra, we solve for the total thrust $4T$:

Equation:

$$4T = ma + f.$$

Substituting known values yields

Equation:

$$4T = ma + f = (2100 \text{ kg})(49 \text{ m/s}^2) + 650 \text{ N}.$$

So the total thrust is

Equation:

$$4T = 1.0 \times 10^5 \text{ N},$$

and the individual thrusts are

Equation:

$$T = \frac{1.0 \times 10^5 \text{ N}}{4} = 2.6 \times 10^4 \text{ N}.$$

Discussion

The numbers are quite large, so the result might surprise you. Experiments such as this were performed in the early 1960s to test the limits of human endurance and the setup designed to protect human subjects in jet fighter emergency ejections. Speeds of 1000 km/h were obtained, with accelerations of 45 *g*'s. (Recall that *g*, the acceleration due to gravity, is 9.80 m/s². When we say that an acceleration is 45 *g*'s, it is 45 × 9.80 m/s², which is approximately 440 m/s².) While living subjects are not used any more, land speeds of 10,000 km/h have been obtained with rocket sleds. In this example, as in the preceding one, the system of interest is obvious. We will see in later examples that choosing the system of interest is crucial—and the choice is not always obvious.

Newton's second law of motion is more than a definition; it is a relationship among acceleration, force, and mass. It can help us make predictions. Each of those physical quantities can be defined independently, so the second law tells us something basic and universal about nature. The next section introduces the third and final law of motion.

Section Summary

- Acceleration, **a**, is defined as a change in velocity, meaning a change in its magnitude or direction, or both.
- An external force is one acting on a system from outside the system, as opposed to internal forces, which act between components within the

system.

- Newton's second law of motion states that the acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system, and inversely proportional to its mass.
- In equation form, Newton's second law of motion is $\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m}$.
- This is often written in the more familiar form: $\mathbf{F}_{\text{net}} = m\mathbf{a}$.
- The weight \mathbf{w} of an object is defined as the force of gravity acting on an object of mass m . The object experiences an acceleration due to gravity \mathbf{g} :

Equation:

$$\mathbf{w} = m\mathbf{g}.$$

- If the only force acting on an object is due to gravity, the object is in free fall.
- Friction is a force that opposes the motion past each other of objects that are touching.

Conceptual Questions

Exercise:

Problem:

Which statement is correct? (a) Net force causes motion. (b) Net force causes change in motion. Explain your answer and give an example.

Exercise:

Problem:

Why can we neglect forces such as those holding a body together when we apply Newton's second law of motion?

Exercise:

Problem:

Explain how the choice of the “system of interest” affects which forces must be considered when applying Newton’s second law of motion.

Exercise:**Problem:**

Describe a situation in which the net external force on a system is not zero, yet its speed remains constant.

Exercise:**Problem:**

A system can have a nonzero velocity while the net external force on it is zero. Describe such a situation.

Exercise:**Problem:**

A rock is thrown straight up. What is the net external force acting on the rock when it is at the top of its trajectory?

Exercise:**Problem:**

(a) Give an example of different net external forces acting on the same system to produce different accelerations. (b) Give an example of the same net external force acting on systems of different masses, producing different accelerations. (c) What law accurately describes both effects? State it in words and as an equation.

Exercise:**Problem:**

If the acceleration of a system is zero, are no external forces acting on it? What about internal forces? Explain your answers.

Exercise:

Problem:

If a constant, nonzero force is applied to an object, what can you say about the velocity and acceleration of the object?

Exercise:**Problem:**

The gravitational force on the basketball in [\[link\]](#) is ignored. When gravity *is* taken into account, what is the direction of the net external force on the basketball—above horizontal, below horizontal, or still horizontal?

Problem Exercises

You may assume data taken from illustrations is accurate to three digits.

Exercise:**Problem:**

A 63.0-kg sprinter starts a race with an acceleration of 4.20 m/s^2 . What is the net external force on him?

Solution:

265 N

Exercise:**Problem:**

If the sprinter from the previous problem accelerates at that rate for 20 m, and then maintains that velocity for the remainder of the 100-m dash, what will be his time for the race?

Exercise:

Problem:

A cleaner pushes a 4.50-kg laundry cart in such a way that the net external force on it is 60.0 N. Calculate the magnitude of its acceleration.

Solution:

$$13.3 \text{ m/s}^2$$

Exercise:**Problem:**

Since astronauts in orbit are apparently weightless, a clever method of measuring their masses is needed to monitor their mass gains or losses to adjust diets. One way to do this is to exert a known force on an astronaut and measure the acceleration produced. Suppose a net external force of 50.0 N is exerted and the astronaut's acceleration is measured to be 0.893 m/s^2 . (a) Calculate her mass. (b) By exerting a force on the astronaut, the vehicle in which they orbit experiences an equal and opposite force. Discuss how this would affect the measurement of the astronaut's acceleration. Propose a method in which recoil of the vehicle is avoided.

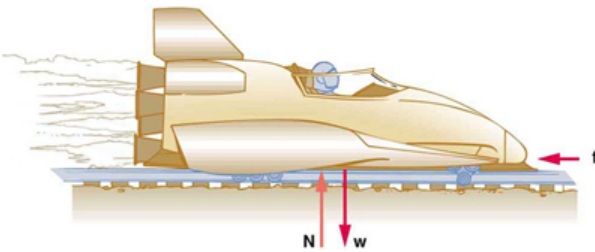
Exercise:**Problem:**

In [\[link\]](#), the net external force on the 24-kg mower is stated to be 51 N. If the force of friction opposing the motion is 24 N, what force F (in newtons) is the person exerting on the mower? Suppose the mower is moving at 1.5 m/s when the force F is removed. How far will the mower go before stopping?

Exercise:

Problem:

The same rocket sled drawn in [\[link\]](#) is decelerated at a rate of 196 m/s^2 . What force is necessary to produce this deceleration? Assume that the rockets are off. The mass of the system is 2100 kg.

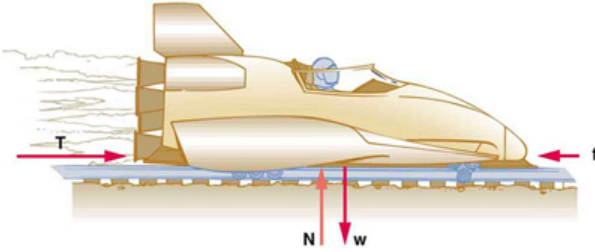
**Exercise:****Problem:**

(a) If the rocket sled shown in [\[link\]](#) starts with only one rocket burning, what is the magnitude of its acceleration? Assume that the mass of the system is 2100 kg, the thrust T is $2.4 \times 10^4 \text{ N}$, and the force of friction opposing the motion is known to be 650 N. (b) Why is the acceleration not one-fourth of what it is with all rockets burning?

Solution:

(a) 12 m/s^2 .

(b) The acceleration is not one-fourth of what it was with all rockets burning because the frictional force is still as large as it was with all rockets burning.



Exercise:

Problem:

What is the deceleration of the rocket sled if it comes to rest in 1.1 s from a speed of 1000 km/h? (Such deceleration caused one test subject to black out and have temporary blindness.)

Exercise:

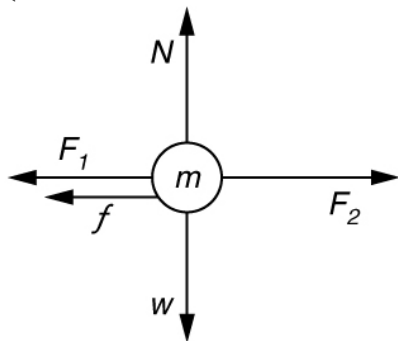
Problem:

Suppose two children push horizontally, but in exactly opposite directions, on a third child in a wagon. The first child exerts a force of 75.0 N, the second a force of 90.0 N, friction is 12.0 N, and the mass of the third child plus wagon is 23.0 kg. (a) What is the system of interest if the acceleration of the child in the wagon is to be calculated? (b) Draw a free-body diagram, including all forces acting on the system. (c) Calculate the acceleration. (d) What would the acceleration be if friction were 15.0 N?

Solution:

(a) The system is the child in the wagon plus the wagon.

(b)



(c) $a = 0.130 \text{ m/s}^2$ in the direction of the second child's push.

(d) $a = 0.00 \text{ m/s}^2$

Exercise:

Problem:

A powerful motorcycle can produce an acceleration of 3.50 m/s^2 while traveling at 90.0 km/h . At that speed the forces resisting motion, including friction and air resistance, total 400 N . (Air resistance is analogous to air friction. It always opposes the motion of an object.) What is the magnitude of the force the motorcycle exerts backward on the ground to produce its acceleration if the mass of the motorcycle with rider is 245 kg ?

Exercise:

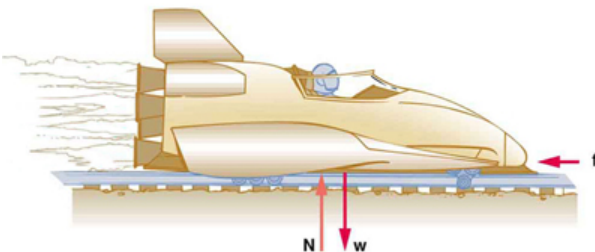
Problem:

The rocket sled shown in [\[link\]](#) accelerates at a rate of 49.0 m/s^2 . Its passenger has a mass of 75.0 kg . (a) Calculate the horizontal component of the force the seat exerts against his body. Compare this with his weight by using a ratio. (b) Calculate the direction and magnitude of the total force the seat exerts against his body.

Solution:

(a) $3.68 \times 10^3 \text{ N}$. This force is 5.00 times greater than his weight.

(b) 3750 N ; 11.3° above horizontal



Exercise:**Problem:**

Repeat the previous problem for the situation in which the rocket sled decelerates at a rate of 201 m/s^2 . In this problem, the forces are exerted by the seat and restraining belts.

Exercise:**Problem:**

The weight of an astronaut plus his space suit on the Moon is only 250 N. How much do they weigh on Earth? What is the mass on the Moon? On Earth?

Solution:

$1.5 \times 10^3 \text{ N}$, 150 kg, 150 kg

Exercise:**Problem:**

Suppose the mass of a fully loaded module in which astronauts take off from the Moon is 10,000 kg. The thrust of its engines is 30,000 N. (a) Calculate its the magnitude of acceleration in a vertical takeoff from the Moon. (b) Could it lift off from Earth? If not, why not? If it could, calculate the magnitude of its acceleration.

Glossary

acceleration

the rate at which an object's velocity changes over a period of time

free-fall

a situation in which the only force acting on an object is the force due to gravity

friction

a force past each other of objects that are touching; examples include rough surfaces and air resistance

net external force

the vector sum of all external forces acting on an object or system; causes a mass to accelerate

Newton's second law of motion

the net external force \mathbf{F}_{net} on an object with mass m is proportional to and in the same direction as the acceleration of the object, \mathbf{a} , and inversely proportional to the mass; defined mathematically as

$$\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m}$$

system

defined by the boundaries of an object or collection of objects being observed; all forces originating from outside of the system are considered external forces

weight

the force \mathbf{w} due to gravity acting on an object of mass m ; defined mathematically as: $\mathbf{w} = m\mathbf{g}$, where \mathbf{g} is the magnitude and direction of the acceleration due to gravity

Newton's Third Law of Motion: Symmetry in Forces

- Understand Newton's third law of motion.
- Apply Newton's third law to define systems and solve problems of motion.

There is a passage in the musical *Man of la Mancha* that relates to Newton's third law of motion. Sancho, in describing a fight with his wife to Don Quixote, says, "Of course I hit her back, Your Grace, but she's a lot harder than me and you know what they say, 'Whether the stone hits the pitcher or the pitcher hits the stone, it's going to be bad for the pitcher.'" This is exactly what happens whenever one body exerts a force on another—the first also experiences a force (equal in magnitude and opposite in direction). Numerous common experiences, such as stubbing a toe or throwing a ball, confirm this. It is precisely stated in **Newton's third law of motion**.

Note:

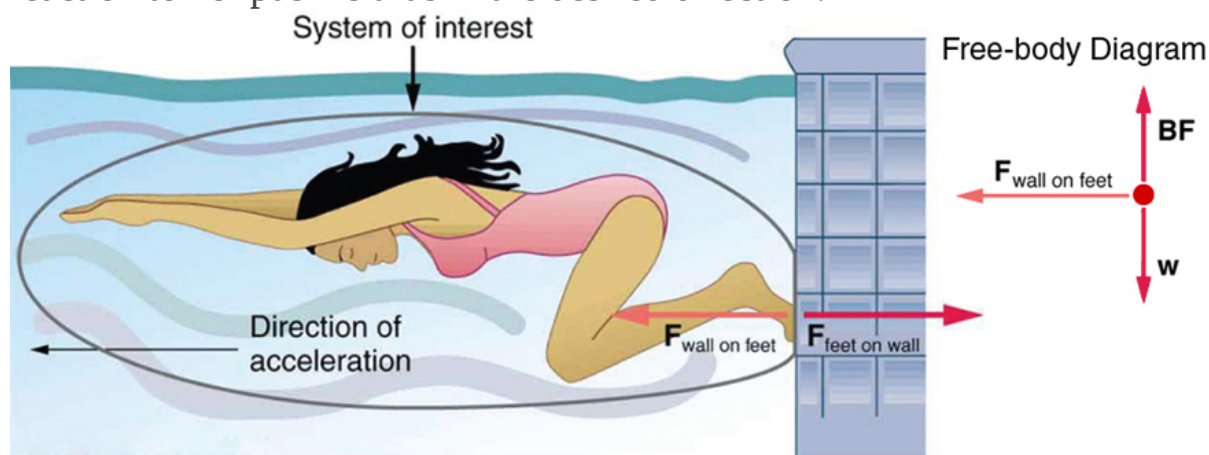
Newton's Third Law of Motion

Whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts.

This law represents a certain *symmetry in nature*: Forces always occur in pairs, and one body cannot exert a force on another without experiencing a force itself. We sometimes refer to this law loosely as "action-reaction," where the force exerted is the action and the force experienced as a consequence is the reaction. Newton's third law has practical uses in analyzing the origin of forces and understanding which forces are external to a system.

We can readily see Newton's third law at work by taking a look at how people move about. Consider a swimmer pushing off from the side of a pool, as illustrated in [\[link\]](#). She pushes against the pool wall with her feet

and accelerates in the direction *opposite* to that of her push. The wall has exerted an equal and opposite force back on the swimmer. You might think that two equal and opposite forces would cancel, but they do not *because they act on different systems*. In this case, there are two systems that we could investigate: the swimmer or the wall. If we select the swimmer to be the system of interest, as in the figure, then $\mathbf{F}_{\text{wall on feet}}$ is an external force on this system and affects its motion. The swimmer moves in the direction of $\mathbf{F}_{\text{wall on feet}}$. In contrast, the force $\mathbf{F}_{\text{feet on wall}}$ acts on the wall and not on our system of interest. Thus $\mathbf{F}_{\text{feet on wall}}$ does not directly affect the motion of the system and does not cancel $\mathbf{F}_{\text{wall on feet}}$. Note that the swimmer pushes in the direction opposite to that in which she wishes to move. The reaction to her push is thus in the desired direction.

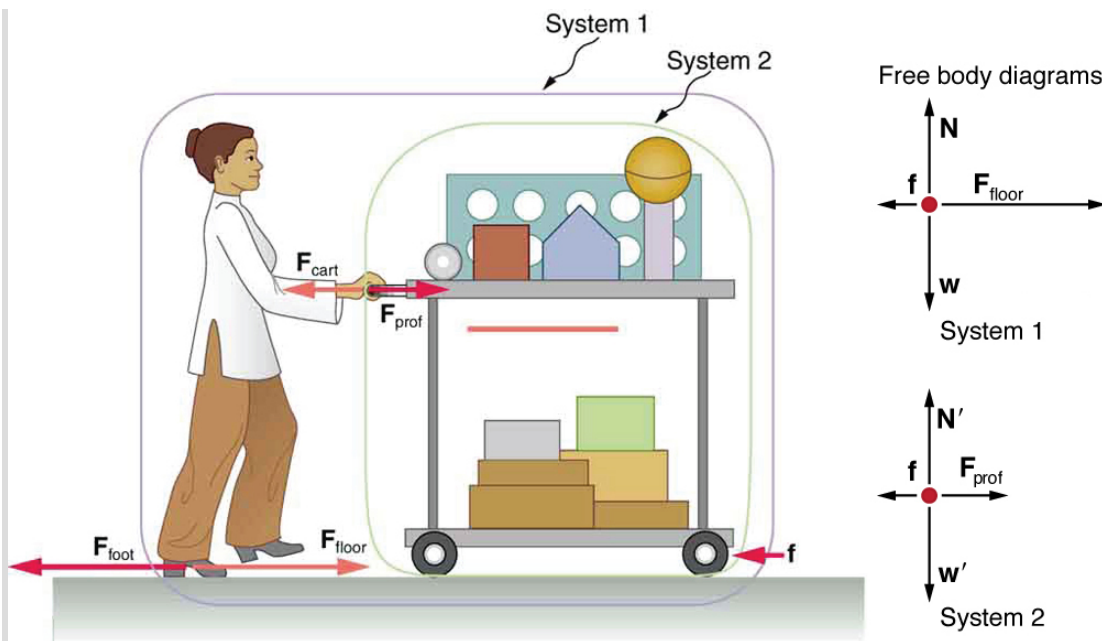


When the swimmer exerts a force $\mathbf{F}_{\text{feet on wall}}$ on the wall, she accelerates in the direction opposite to that of her push. This means the net external force on her is in the direction opposite to $\mathbf{F}_{\text{feet on wall}}$. This opposition occurs because, in accordance with Newton's third law of motion, the wall exerts a force $\mathbf{F}_{\text{wall on feet}}$ on her, equal in magnitude but in the direction opposite to the one she exerts on it. The line around the swimmer indicates the system of interest. Note that $\mathbf{F}_{\text{feet on wall}}$ does not act on this system (the swimmer) and, thus, does not cancel $\mathbf{F}_{\text{wall on feet}}$. Thus the free-body diagram shows only $\mathbf{F}_{\text{wall on feet}}$, \mathbf{w} , the gravitational force, and \mathbf{BF} , the buoyant force of the water supporting the swimmer's weight. The vertical forces \mathbf{w} and \mathbf{BF} cancel since there is no vertical motion.

Other examples of Newton's third law are easy to find. As a professor paces in front of a whiteboard, she exerts a force backward on the floor. The floor exerts a reaction force forward on the professor that causes her to accelerate forward. Similarly, a car accelerates because the ground pushes forward on the drive wheels in reaction to the drive wheels pushing backward on the ground. You can see evidence of the wheels pushing backward when tires spin on a gravel road and throw rocks backward. In another example, rockets move forward by expelling gas backward at high velocity. This means the rocket exerts a large backward force on the gas in the rocket combustion chamber, and the gas therefore exerts a large reaction force forward on the rocket. This reaction force is called **thrust**. It is a common misconception that rockets propel themselves by pushing on the ground or on the air behind them. They actually work better in a vacuum, where they can more readily expel the exhaust gases. Helicopters similarly create lift by pushing air down, thereby experiencing an upward reaction force. Birds and airplanes also fly by exerting force on air in a direction opposite to that of whatever force they need. For example, the wings of a bird force air downward and backward in order to get lift and move forward. An octopus propels itself in the water by ejecting water through a funnel from its body, similar to a jet ski. In a situation similar to Sancho's, professional cage fighters experience reaction forces when they punch, sometimes breaking their hand by hitting an opponent's body.

Example:**Getting Up To Speed: Choosing the Correct System**

A physics professor pushes a cart of demonstration equipment to a lecture hall, as seen in [\[link\]](#). Her mass is 65.0 kg, the cart's is 12.0 kg, and the equipment's is 7.0 kg. Calculate the acceleration produced when the professor exerts a backward force of 150 N on the floor. All forces opposing the motion, such as friction on the cart's wheels and air resistance, total 24.0 N.



A professor pushes a cart of demonstration equipment. The lengths of the arrows are proportional to the magnitudes of the forces (except for f , since it is too small to draw to scale). Different questions are asked in each example; thus, the system of interest must be defined differently for each. System 1 is appropriate for this example, since it asks for the acceleration of the entire group of objects. Only F_{floor} and f are external forces acting on System 1 along the line of motion. All other forces either cancel or act on the outside world. System 2 is chosen for [\[link\]](#) so that F_{prof} will be an external force and enter into Newton's second law. Note that the free-body diagrams, which allow us to apply Newton's second law, vary with the system chosen.

Strategy

Since they accelerate as a unit, we define the system to be the professor, cart, and equipment. This is System 1 in [\[link\]](#). The professor pushes backward with a force F_{foot} of 150 N. According to Newton's third law, the floor exerts a forward reaction force F_{floor} of 150 N on System 1. Because all motion is horizontal, we can assume there is no net force in the vertical direction. The problem is therefore one-dimensional along the

horizontal direction. As noted, \mathbf{f} opposes the motion and is thus in the opposite direction of $\mathbf{F}_{\text{floor}}$. Note that we do not include the forces \mathbf{F}_{prof} or \mathbf{F}_{cart} because these are internal forces, and we do not include \mathbf{F}_{foot} because it acts on the floor, not on the system. There are no other significant forces acting on System 1. If the net external force can be found from all this information, we can use Newton's second law to find the acceleration as requested. See the free-body diagram in the figure.

Solution

Newton's second law is given by

Equation:

$$a = \frac{F_{\text{net}}}{m}.$$

The net external force on System 1 is deduced from [\[link\]](#) and the discussion above to be

Equation:

$$F_{\text{net}} = F_{\text{floor}} - f = 150 \text{ N} - 24.0 \text{ N} = 126 \text{ N}.$$

The mass of System 1 is

Equation:

$$m = (65.0 + 12.0 + 7.0) \text{ kg} = 84 \text{ kg}.$$

These values of F_{net} and m produce an acceleration of

Equation:

$$a = \frac{F_{\text{net}}}{m},$$
$$a = \frac{126 \text{ N}}{84 \text{ kg}} = 1.5 \text{ m/s}^2.$$

Discussion

None of the forces between components of System 1, such as between the professor's hands and the cart, contribute to the net external force because they are internal to System 1. Another way to look at this is to note that forces between components of a system cancel because they are equal in magnitude and opposite in direction. For example, the force exerted by the

professor on the cart results in an equal and opposite force back on her. In this case both forces act on the same system and, therefore, cancel. Thus internal forces (between components of a system) cancel. Choosing System 1 was crucial to solving this problem.

Example:

Force on the Cart—Choosing a New System

Calculate the force the professor exerts on the cart in [\[link\]](#) using data from the previous example if needed.

Strategy

If we now define the system of interest to be the cart plus equipment (System 2 in [\[link\]](#)), then the net external force on System 2 is the force the professor exerts on the cart minus friction. The force she exerts on the cart, \mathbf{F}_{prof} , is an external force acting on System 2. \mathbf{F}_{prof} was internal to System 1, but it is external to System 2 and will enter Newton's second law for System 2.

Solution

Newton's second law can be used to find \mathbf{F}_{prof} . Starting with

Equation:

$$a = \frac{F_{\text{net}}}{m}$$

and noting that the magnitude of the net external force on System 2 is

Equation:

$$F_{\text{net}} = F_{\text{prof}} - f,$$

we solve for F_{prof} , the desired quantity:

Equation:

$$F_{\text{prof}} = F_{\text{net}} + f.$$

The value of f is given, so we must calculate net F_{net} . That can be done since both the acceleration and mass of System 2 are known. Using Newton's second law we see that

Equation:

$$F_{\text{net}} = ma,$$

where the mass of System 2 is 19.0 kg ($m = 12.0 \text{ kg} + 7.0 \text{ kg}$) and its acceleration was found to be $a = 1.5 \text{ m/s}^2$ in the previous example. Thus,

Equation:

$$F_{\text{net}} = ma,$$

Equation:

$$F_{\text{net}} = (19.0 \text{ kg})(1.5 \text{ m/s}^2) = 29 \text{ N}.$$

Now we can find the desired force:

Equation:

$$F_{\text{prof}} = F_{\text{net}} + f,$$

Equation:

$$F_{\text{prof}} = 29 \text{ N} + 24.0 \text{ N} = 53 \text{ N}.$$

Discussion

It is interesting that this force is significantly less than the 150-N force the professor exerted backward on the floor. Not all of that 150-N force is transmitted to the cart; some of it accelerates the professor.

The choice of a system is an important analytical step both in solving problems and in thoroughly understanding the physics of the situation (which is not necessarily the same thing).

Note:

PhET Explorations: Gravity Force Lab

Visualize the gravitational force that two objects exert on each other.

Change properties of the objects in order to see how it changes the gravity force.

https://phet.colorado.edu/sims/html/gravity-force-lab/latest/gravity-force-lab_en.html

Section Summary

- **Newton's third law of motion** represents a basic symmetry in nature. It states: Whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that the first body exerts.
- A **thrust** is a reaction force that pushes a body forward in response to a backward force. Rockets, airplanes, and cars are pushed forward by a thrust reaction force.

Conceptual Questions

Exercise:

Problem:

When you take off in a jet aircraft, there is a sensation of being pushed back into the seat. Explain why you move backward in the seat—is there really a force backward on you? (The same reasoning explains whiplash injuries, in which the head is apparently thrown backward.)

Exercise:

Problem:

A device used since the 1940s to measure the kick or recoil of the body due to heart beats is the “ballistocardiograph.” What physics principle(s) are involved here to measure the force of cardiac contraction? How might we construct such a device?

Exercise:

Problem:

Describe a situation in which one system exerts a force on another and, as a consequence, experiences a force that is equal in magnitude and opposite in direction. Which of Newton's laws of motion apply?

Exercise:**Problem:**

Why does an ordinary rifle recoil (kick backward) when fired? The barrel of a recoilless rifle is open at both ends. Describe how Newton's third law applies when one is fired. Can you safely stand close behind one when it is fired?

Exercise:**Problem:**

An American football lineman reasons that it is senseless to try to out-push the opposing player, since no matter how hard he pushes he will experience an equal and opposite force from the other player. Use Newton's laws and draw a free-body diagram of an appropriate system to explain how he can still out-push the opposition if he is strong enough.

Exercise:**Problem:**

Newton's third law of motion tells us that forces always occur in pairs of equal and opposite magnitude. Explain how the choice of the "system of interest" affects whether one such pair of forces cancels.

Problem Exercises**Exercise:**

Problem:

What net external force is exerted on a 1100-kg artillery shell fired from a battleship if the shell is accelerated at $2.40 \times 10^4 \text{ m/s}^2$? What is the magnitude of the force exerted on the ship by the artillery shell?

Solution:

Force on shell: $2.64 \times 10^7 \text{ N}$

Force exerted on ship = $-2.64 \times 10^7 \text{ N}$, by Newton's third law

Exercise:**Problem:**

A brave but inadequate rugby player is being pushed backward by an opposing player who is exerting a force of 800 N on him. The mass of the losing player plus equipment is 90.0 kg, and he is accelerating at 1.20 m/s^2 backward. (a) What is the force of friction between the losing player's feet and the grass? (b) What force does the winning player exert on the ground to move forward if his mass plus equipment is 110 kg? (c) Draw a sketch of the situation showing the system of interest used to solve each part. For this situation, draw a free-body diagram and write the net force equation.

Glossary**Newton's third law of motion**

whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that the first body exerts

thrust

a reaction force that pushes a body forward in response to a backward force; rockets, airplanes, and cars are pushed forward by a thrust reaction force

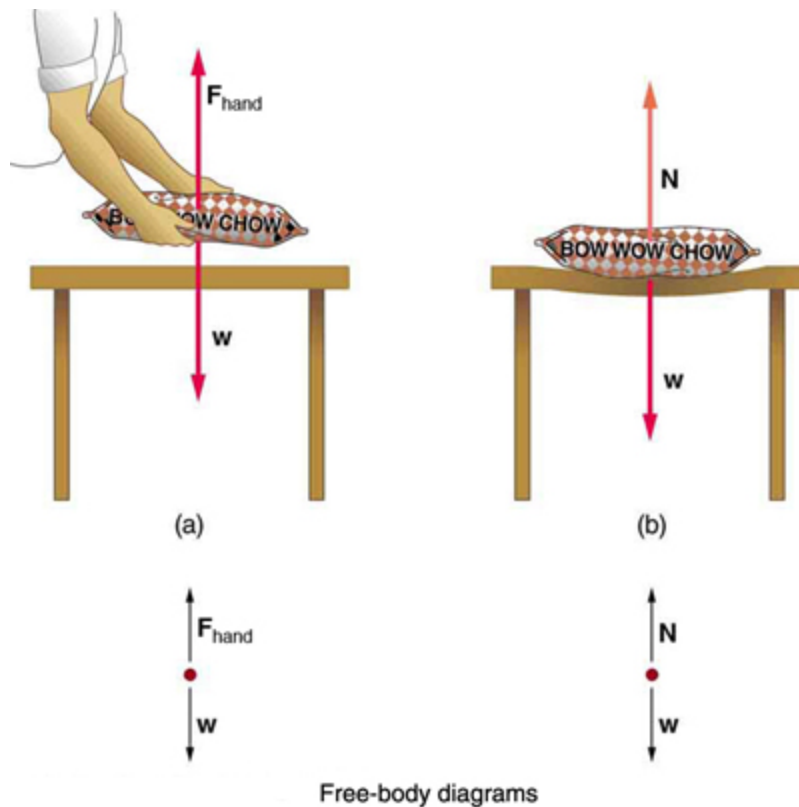
Normal, Tension, and Other Examples of Forces

- Define normal and tension forces.
- Apply Newton's laws of motion to solve problems involving a variety of forces.
- Use trigonometric identities to resolve weight into components.

Forces are given many names, such as push, pull, thrust, lift, weight, friction, and tension. Traditionally, forces have been grouped into several categories and given names relating to their source, how they are transmitted, or their effects. The most important of these categories are discussed in this section, together with some interesting applications. Further examples of forces are discussed later in this text.

Normal Force

Weight (also called force of gravity) is a pervasive force that acts at all times and must be counteracted to keep an object from falling. You definitely notice that you must support the weight of a heavy object by pushing up on it when you hold it stationary, as illustrated in [\[link\]](#)(a). But how do inanimate objects like a table support the weight of a mass placed on them, such as shown in [\[link\]](#)(b)? When the bag of dog food is placed on the table, the table actually sags slightly under the load. This would be noticeable if the load were placed on a card table, but even rigid objects deform when a force is applied to them. Unless the object is deformed beyond its limit, it will exert a restoring force much like a deformed spring (or trampoline or diving board). The greater the deformation, the greater the restoring force. So when the load is placed on the table, the table sags until the restoring force becomes as large as the weight of the load. At this point the net external force on the load is zero. That is the situation when the load is stationary on the table. The table sags quickly, and the sag is slight so we do not notice it. But it is similar to the sagging of a trampoline when you climb onto it.



(a) The person holding the bag of dog food must supply an upward force \mathbf{F}_{hand} equal in magnitude and opposite in direction to the weight of the food \mathbf{w} . (b) The card table sags when the dog food is placed on it, much like a stiff trampoline. Elastic restoring forces in the table grow as it sags until they supply a force \mathbf{N} equal in magnitude and opposite in direction to the weight of the load.

We must conclude that whatever supports a load, be it animate or not, must supply an upward force equal to the weight of the load, as we assumed in a few of the previous examples. If the force supporting a load is perpendicular to the surface of contact between the load and its support, this force is defined to be a **normal force** and here is given the symbol \mathbf{N} . (This is not the unit for force N.) The word *normal* means perpendicular to a

surface. The normal force can be less than the object's weight if the object is on an incline, as you will see in the next example.

Note:

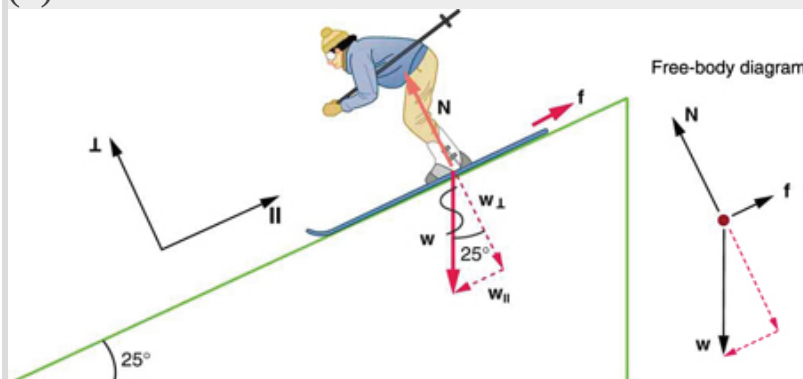
Common Misconception: Normal Force (\mathbf{N}) vs. Newton (N)

In this section we have introduced the quantity normal force, which is represented by the variable \mathbf{N} . This should not be confused with the symbol for the newton, which is also represented by the letter N . These symbols are particularly important to distinguish because the units of a normal force (\mathbf{N}) happen to be newtons (N). For example, the normal force \mathbf{N} that the floor exerts on a chair might be $\mathbf{N} = 100 \text{ N}$. One important difference is that normal force is a vector, while the newton is simply a unit. Be careful not to confuse these letters in your calculations! You will encounter more similarities among variables and units as you proceed in physics. Another example of this is the quantity work (W) and the unit watts (W).

Example:

Weight on an Incline, a Two-Dimensional Problem

Consider the skier on a slope shown in [\[link\]](#). Her mass including equipment is 60.0 kg . (a) What is her acceleration if friction is negligible? (b) What is her acceleration if friction is known to be 45.0 N ?



Since motion and friction are parallel to the slope, it is most convenient to project all

forces onto a coordinate system where one axis is parallel to the slope and the other is perpendicular (axes shown to left of skier).

\mathbf{N} is perpendicular to the slope and \mathbf{f} is parallel to the slope, but \mathbf{w} has components along both axes, namely \mathbf{w}_\perp and \mathbf{w}_\parallel . \mathbf{N} is equal in magnitude to \mathbf{w}_\perp , so that there is no motion perpendicular to the slope, but f is less than w_\parallel , so that there is a downslope acceleration (along the parallel axis).

Strategy

This is a two-dimensional problem, since the forces on the skier (the system of interest) are not parallel. The approach we have used in two-dimensional kinematics also works very well here. Choose a convenient coordinate system and project the vectors onto its axes, creating two connected *one*-dimensional problems to solve. The most convenient coordinate system for motion on an incline is one that has one coordinate parallel to the slope and one perpendicular to the slope. (Remember that motions along mutually perpendicular axes are independent.) We use the symbols \perp and \parallel to represent perpendicular and parallel, respectively. This choice of axes simplifies this type of problem, because there is no motion perpendicular to the slope and because friction is always parallel to the surface between two objects. The only external forces acting on the system are the skier's weight, friction, and the support of the slope, respectively labeled \mathbf{w} , \mathbf{f} , and \mathbf{N} in [\[link\]](#). \mathbf{N} is always perpendicular to the slope, and \mathbf{f} is parallel to it. But \mathbf{w} is not in the direction of either axis, and so the first step we take is to project it into components along the chosen axes, defining w_\parallel to be the component of weight parallel to the slope and w_\perp the component of weight perpendicular to the slope. Once this is done, we can consider the two separate problems of forces parallel to the slope and forces perpendicular to the slope.

Solution

The magnitude of the component of the weight parallel to the slope is $w_\parallel = w \sin(25^\circ) = mg \sin(25^\circ)$, and the magnitude of the component of

the weight perpendicular to the slope is

$$w_{\perp} = w \cos (25^{\circ}) = mg \cos (25^{\circ}).$$

(a) Neglecting friction. Since the acceleration is parallel to the slope, we need only consider forces parallel to the slope. (Forces perpendicular to the slope add to zero, since there is no acceleration in that direction.) The forces parallel to the slope are the amount of the skier's weight parallel to the slope w_{\parallel} and friction f . Using Newton's second law, with subscripts to denote quantities parallel to the slope,

Equation:

$$a_{\parallel} = \frac{F_{\text{net}\parallel}}{m}$$

where $F_{\text{net}\parallel} = w_{\parallel} = mg \sin (25^{\circ})$, assuming no friction for this part, so that

Equation:

$$a_{\parallel} = \frac{F_{\text{net}\parallel}}{m} = \frac{mg \sin (25^{\circ})}{m} = g \sin (25^{\circ})$$

Equation:

$$(9.80 \text{ m/s}^2)(0.4226) = 4.14 \text{ m/s}^2$$

is the acceleration.

(b) Including friction. We now have a given value for friction, and we know its direction is parallel to the slope and it opposes motion between surfaces in contact. So the net external force is now

Equation:

$$F_{\text{net}\parallel} = w_{\parallel} - f,$$

and substituting this into Newton's second law, $a_{\parallel} = \frac{F_{\text{net}\parallel}}{m}$, gives

Equation:

$$a_{\parallel} = \frac{F_{\text{net}\parallel}}{m} = \frac{w_{\parallel} - f}{m} = \frac{mg \sin (25^{\circ}) - f}{m}.$$

We substitute known values to obtain

Equation:

$$a_{\parallel} = \frac{(60.0 \text{ kg})(9.80 \text{ m/s}^2)(0.4226) - 45.0 \text{ N}}{60.0 \text{ kg}},$$

which yields

Equation:

$$a_{\parallel} = 3.39 \text{ m/s}^2,$$

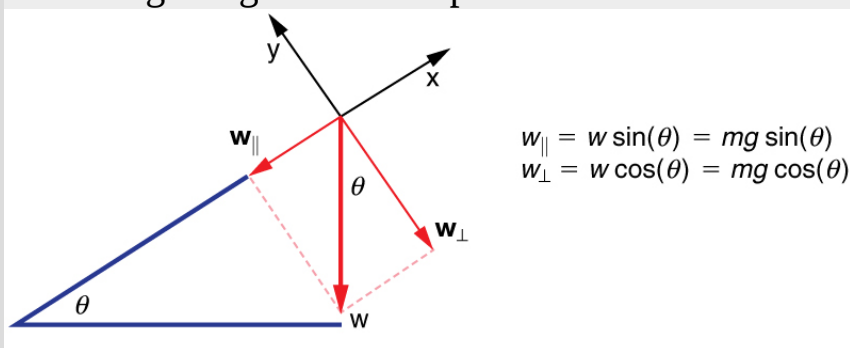
which is the acceleration parallel to the incline when there is 45.0 N of opposing friction.

Discussion

Since friction always opposes motion between surfaces, the acceleration is smaller when there is friction than when there is none. In fact, it is a general result that if friction on an incline is negligible, then the acceleration down the incline is $a = g \sin\theta$, *regardless of mass*. This is related to the previously discussed fact that all objects fall with the same acceleration in the absence of air resistance. Similarly, all objects, regardless of mass, slide down a frictionless incline with the same acceleration (if the angle is the same).

Note:

Resolving Weight into Components



An object rests on an incline that makes an

angle θ with the horizontal.

When an object rests on an incline that makes an angle θ with the horizontal, the force of gravity acting on the object is divided into two components: a force acting perpendicular to the plane, \mathbf{w}_{\perp} , and a force acting parallel to the plane, \mathbf{w}_{\parallel} . The perpendicular force of weight, \mathbf{w}_{\perp} , is typically equal in magnitude and opposite in direction to the normal force, \mathbf{N} . The force acting parallel to the plane, \mathbf{w}_{\parallel} , causes the object to accelerate down the incline. The force of friction, \mathbf{f} , opposes the motion of the object, so it acts upward along the plane.

It is important to be careful when resolving the weight of the object into components. If the angle of the incline is at an angle θ to the horizontal, then the magnitudes of the weight components are

Equation:

$$w_{\parallel} = w \sin (\theta) = mg \sin (\theta)$$

and

Equation:

$$w_{\perp} = w \cos (\theta) = mg \cos (\theta).$$

Instead of memorizing these equations, it is helpful to be able to determine them from reason. To do this, draw the right triangle formed by the three weight vectors. Notice that the angle θ of the incline is the same as the angle formed between \mathbf{w} and \mathbf{w}_{\perp} . Knowing this property, you can use trigonometry to determine the magnitude of the weight components:

Equation:

$$\begin{aligned} \cos (\theta) &= \frac{w_{\perp}}{w} \\ w_{\perp} &= w \cos (\theta) = mg \cos (\theta) \end{aligned}$$

Equation:

$$\begin{aligned} \sin (\theta) &= \frac{w_{\parallel}}{w} \\ w_{\parallel} &= w \sin (\theta) = mg \sin (\theta) \end{aligned}$$

Note:

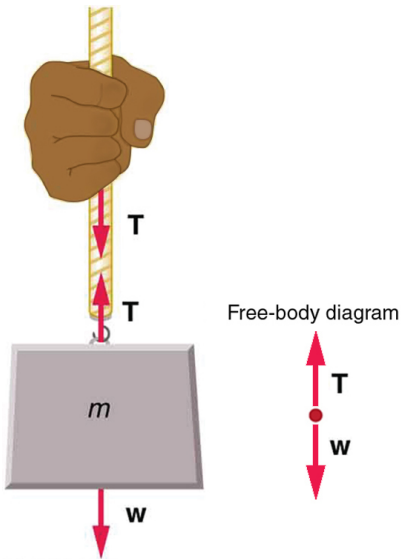
Take-Home Experiment: Force Parallel

To investigate how a force parallel to an inclined plane changes, find a rubber band, some objects to hang from the end of the rubber band, and a board you can position at different angles. How much does the rubber band stretch when you hang the object from the end of the board? Now place the board at an angle so that the object slides off when placed on the board. How much does the rubber band extend if it is lined up parallel to the board and used to hold the object stationary on the board? Try two more angles. What does this show?

Tension

A **tension** is a force along the length of a medium, especially a force carried by a flexible medium, such as a rope or cable. The word “tension” comes from a Latin word meaning “to stretch.” Not coincidentally, the flexible cords that carry muscle forces to other parts of the body are called *tendons*. Any flexible connector, such as a string, rope, chain, wire, or cable, can exert pulls only parallel to its length; thus, a force carried by a flexible connector is a tension with direction parallel to the connector. It is important to understand that tension is a pull in a connector. In contrast, consider the phrase: “You can’t push a rope.” The tension force pulls outward along the two ends of a rope.

Consider a person holding a mass on a rope as shown in [\[link\]](#).



When a perfectly flexible connector (one requiring no force to bend it) such as this rope transmits a force \mathbf{T} , that force must be parallel to the length of the rope, as shown. The pull such a flexible connector exerts is a tension. Note that the rope pulls with equal force but in opposite directions on the hand and the supported mass (neglecting the weight of the rope). This is an example of Newton's third law. The rope is the medium that carries

the equal and opposite forces between the two objects. The tension anywhere in the rope between the hand and the mass is equal. Once you have determined the tension in one location, you have determined the tension at all locations along the rope.

Tension in the rope must equal the weight of the supported mass, as we can prove using Newton's second law. If the 5.00-kg mass in the figure is stationary, then its acceleration is zero, and thus $\mathbf{F}_{\text{net}} = 0$. The only external forces acting on the mass are its weight \mathbf{w} and the tension \mathbf{T} supplied by the rope. Thus,

Equation:

$$F_{\text{net}} = T - w = 0,$$

where T and w are the magnitudes of the tension and weight and their signs indicate direction, with up being positive here. Thus, just as you would expect, the tension equals the weight of the supported mass:

Equation:

$$T = w = mg.$$

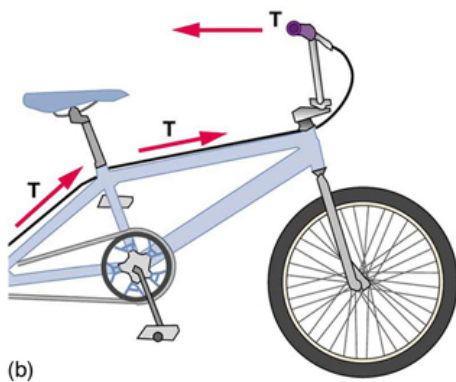
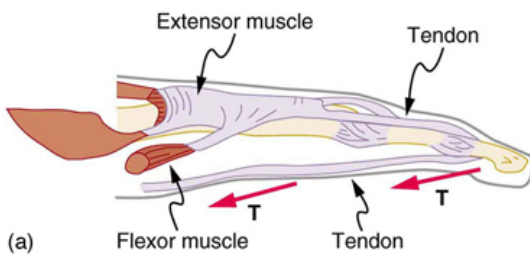
For a 5.00-kg mass, then (neglecting the mass of the rope) we see that

Equation:

$$T = mg = (5.00 \text{ kg})(9.80 \text{ m/s}^2) = 49.0 \text{ N}.$$

If we cut the rope and insert a spring, the spring would extend a length corresponding to a force of 49.0 N, providing a direct observation and measure of the tension force in the rope.

Flexible connectors are often used to transmit forces around corners, such as in a hospital traction system, a finger joint, or a bicycle brake cable. If there is no friction, the tension is transmitted undiminished. Only its direction changes, and it is always parallel to the flexible connector. This is illustrated in [\[link\]](#) (a) and (b).



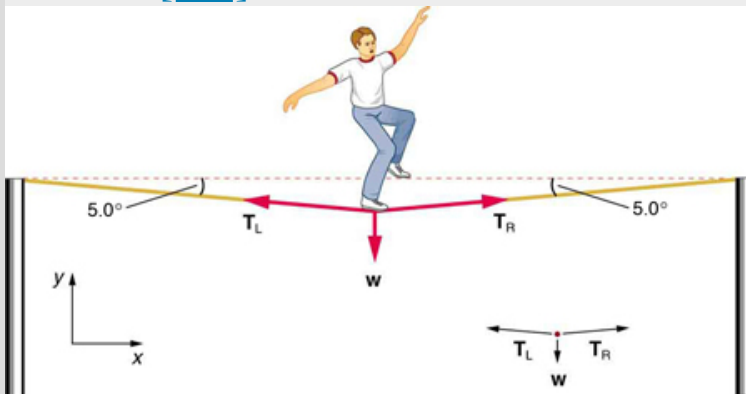
(a) Tendons in the finger carry force \mathbf{T} from the muscles to other parts of the finger, usually changing the force's direction, but not its magnitude (the

tendons are relatively friction free). (b) The brake cable on a bicycle carries the tension \mathbf{T} from the handlebars to the brake mechanism. Again, the direction but not the magnitude of \mathbf{T} is changed.

Example:

What Is the Tension in a Tightrope?

Calculate the tension in the wire supporting the 70.0-kg tightrope walker shown in [\[link\]](#).



The weight of a tightrope walker causes a wire to sag by 5.0 degrees. The system of interest here is the point in the wire at which the tightrope walker is standing.

Strategy

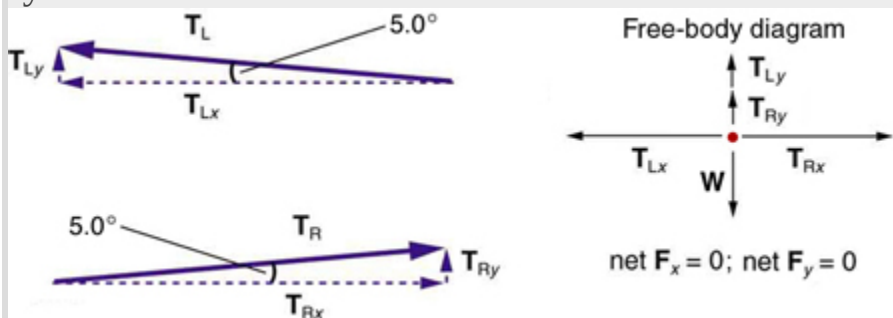
As you can see in the figure, the wire is not perfectly horizontal (it cannot be!), but is bent under the person's weight. Thus, the tension on either side of the person has an upward component that can support his weight. As

usual, forces are vectors represented pictorially by arrows having the same directions as the forces and lengths proportional to their magnitudes. The system is the tightrope walker, and the only external forces acting on him are his weight \mathbf{w} and the two tensions \mathbf{T}_L (left tension) and \mathbf{T}_R (right tension), as illustrated. It is reasonable to neglect the weight of the wire itself. The net external force is zero since the system is stationary. A little trigonometry can now be used to find the tensions. One conclusion is possible at the outset—we can see from part (b) of the figure that the magnitudes of the tensions T_L and T_R must be equal. This is because there is no horizontal acceleration in the rope, and the only forces acting to the left and right are T_L and T_R . Thus, the magnitude of those forces must be equal so that they cancel each other out.

Whenever we have two-dimensional vector problems in which no two vectors are parallel, the easiest method of solution is to pick a convenient coordinate system and project the vectors onto its axes. In this case the best coordinate system has one axis horizontal and the other vertical. We call the horizontal the x -axis and the vertical the y -axis.

Solution

First, we need to resolve the tension vectors into their horizontal and vertical components. It helps to draw a new free-body diagram showing all of the horizontal and vertical components of each force acting on the system.



When the vectors are projected onto vertical and horizontal axes, their components along those axes must add to zero, since the tightrope walker is stationary. The small angle results in T being much greater than w .

Consider the horizontal components of the forces (denoted with a subscript x):

Equation:

$$F_{\text{net}x} = T_{Lx} - T_{Rx}.$$

The net external horizontal force $F_{\text{net}x} = 0$, since the person is stationary. Thus,

Equation:

$$\begin{aligned} F_{\text{net}x} = 0 &= T_{Lx} - T_{Rx} \\ T_{Lx} &= T_{Rx}. \end{aligned}$$

Now, observe [\[link\]](#). You can use trigonometry to determine the magnitude of T_L and T_R . Notice that:

Equation:

$$\begin{aligned} \cos(5.0^\circ) &= \frac{T_{Lx}}{T_L} \\ T_{Lx} &= T_L \cos(5.0^\circ) \\ \cos(5.0^\circ) &= \frac{T_{Rx}}{T_R} \\ T_{Rx} &= T_R \cos(5.0^\circ). \end{aligned}$$

Equating T_{Lx} and T_{Rx} :

Equation:

$$T_L \cos(5.0^\circ) = T_R \cos(5.0^\circ).$$

Thus,

Equation:

$$T_L = T_R = T,$$

as predicted. Now, considering the vertical components (denoted by a subscript y), we can solve for T . Again, since the person is stationary, Newton's second law implies that net $F_y = 0$. Thus, as illustrated in the free-body diagram in [\[link\]](#),

Equation:

$$F_{\text{net}y} = T_{Ly} + T_{Ry} - w = 0.$$

Observing [\[link\]](#), we can use trigonometry to determine the relationship between T_{Ly} , T_{Ry} , and T . As we determined from the analysis in the horizontal direction, $T_L = T_R = T$:

Equation:

$$\begin{aligned}\sin (5.0^\circ) &= \frac{T_{Ly}}{T_L} \\ T_{Ly} = T_L \sin (5.0^\circ) &= T \sin (5.0^\circ) \\ \sin (5.0^\circ) &= \frac{T_{Ry}}{T_R} \\ T_{Ry} = T_R \sin (5.0^\circ) &= T \sin (5.0^\circ).\end{aligned}$$

Now, we can substitute the values for T_{Ly} and T_{Ry} , into the net force equation in the vertical direction:

Equation:

$$\begin{aligned}F_{\text{net}y} &= T_{Ly} + T_{Ry} - w = 0 \\ F_{\text{net}y} &= T \sin (5.0^\circ) + T \sin (5.0^\circ) - w = 0 \\ 2 T \sin (5.0^\circ) - w &= 0 \\ 2 T \sin (5.0^\circ) &= w\end{aligned}$$

and

Equation:

$$T = \frac{w}{2 \sin (5.0^\circ)} = \frac{mg}{2 \sin (5.0^\circ)},$$

so that

Equation:

$$T = \frac{(70.0 \text{ kg})(9.80 \text{ m/s}^2)}{2(0.0872)},$$

and the tension is

Equation:

$$T = 3900 \text{ N}.$$

Discussion

Note that the vertical tension in the wire acts as a normal force that supports the weight of the tightrope walker. The tension is almost six times the 686-N weight of the tightrope walker. Since the wire is nearly horizontal, the vertical component of its tension is only a small fraction of the tension in the wire. The large horizontal components are in opposite directions and cancel, and so most of the tension in the wire is not used to support the weight of the tightrope walker.

If we wish to *create* a very large tension, all we have to do is exert a force perpendicular to a flexible connector, as illustrated in [\[link\]](#). As we saw in the last example, the weight of the tightrope walker acted as a force perpendicular to the rope. We saw that the tension in the rope related to the weight of the tightrope walker in the following way:

Equation:

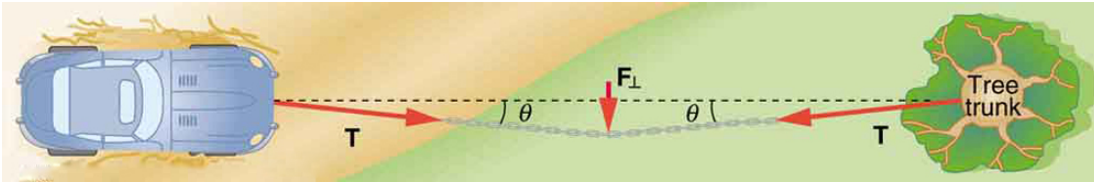
$$T = \frac{w}{2 \sin (\theta)}.$$

We can extend this expression to describe the tension T created when a perpendicular force (\mathbf{F}_\perp) is exerted at the middle of a flexible connector:

Equation:

$$T = \frac{F_\perp}{2 \sin (\theta)}.$$

Note that θ is the angle between the horizontal and the bent connector. In this case, T becomes very large as θ approaches zero. Even the relatively small weight of any flexible connector will cause it to sag, since an infinite tension would result if it were horizontal (i.e., $\theta = 0$ and $\sin \theta = 0$). (See [\[link\]](#).)



We can create a very large tension in the chain by pushing on it perpendicular to its length, as shown. Suppose we wish to pull a car out of the mud when no tow truck is available. Each time the car moves forward, the chain is tightened to keep it as nearly straight as possible. The tension in the chain is given by $T = \frac{F_{\perp}}{2 \sin(\theta)}$; since θ is small, T is very large. This situation is analogous to the tightrope walker shown in [\[link\]](#), except that the tensions shown here are those transmitted to the car and the tree rather than those acting at the point where \mathbf{F}_{\perp} is applied.



Unless an infinite tension is exerted, any flexible connector—such as the chain at the bottom of the picture—will sag under its own weight, giving a characteristic curve when the weight is evenly

distributed along the length.

Suspension bridges—such as the Golden Gate Bridge shown in this image—are essentially very heavy flexible connectors. The weight of the bridge is evenly distributed along the length of flexible connectors, usually cables, which take on the characteristic shape.

(credit: Leaflet, Wikimedia Commons)

Extended Topic: Real Forces and Inertial Frames

There is another distinction among forces in addition to the types already mentioned. Some forces are real, whereas others are not. *Real forces* are those that have some physical origin, such as the gravitational pull. Contrastingly, *fictitious forces* are those that arise simply because an observer is in an accelerating frame of reference, such as one that rotates (like a merry-go-round) or undergoes linear acceleration (like a car slowing down). For example, if a satellite is heading due north above Earth's northern hemisphere, then to an observer on Earth it will appear to experience a force to the west that has no physical origin. Of course, what is happening here is that Earth is rotating toward the east and moves east under the satellite. In Earth's frame this looks like a westward force on the satellite, or it can be interpreted as a violation of Newton's first law (the law of inertia). An **inertial frame of reference** is one in which all forces are real and, equivalently, one in which Newton's laws have the simple forms given in this chapter.

Earth's rotation is slow enough that Earth is nearly an inertial frame. You ordinarily must perform precise experiments to observe fictitious forces and the slight departures from Newton's laws, such as the effect just described. On the large scale, such as for the rotation of weather systems and ocean currents, the effects can be easily observed.

The crucial factor in determining whether a frame of reference is inertial is whether it accelerates or rotates relative to a known inertial frame. Unless stated otherwise, all phenomena discussed in this text are considered in inertial frames.

All the forces discussed in this section are real forces, but there are a number of other real forces, such as lift and thrust, that are not discussed in this section. They are more specialized, and it is not necessary to discuss every type of force. It is natural, however, to ask where the basic simplicity we seek to find in physics is in the long list of forces. Are some more basic than others? Are some different manifestations of the same underlying force? The answer to both questions is yes, as will be seen in the next (extended) section and in the treatment of modern physics later in the text.

Note:**PhET Explorations: Forces in 1 Dimension**

Explore the forces at work when you try to push a filing cabinet. Create an applied force and see the resulting friction force and total force acting on the cabinet. Charts show the forces, position, velocity, and acceleration vs. time. View a free-body diagram of all the forces (including gravitational and normal forces).

[Forces in
1
Dimension](#)

Section Summary

- When objects rest on a surface, the surface applies a force to the object that supports the weight of the object. This supporting force acts

perpendicular to and away from the surface. It is called a normal force, **N**.

- When objects rest on a non-accelerating horizontal surface, the magnitude of the normal force is equal to the weight of the object:

Equation:

$$N = mg.$$

- When objects rest on an inclined plane that makes an angle θ with the horizontal surface, the weight of the object can be resolved into components that act perpendicular (\mathbf{w}_\perp) and parallel (\mathbf{w}_\parallel) to the surface of the plane. These components can be calculated using:

Equation:

$$w_\parallel = w \sin(\theta) = mg \sin(\theta)$$

Equation:

$$w_\perp = w \cos(\theta) = mg \cos(\theta).$$

- The pulling force that acts along a stretched flexible connector, such as a rope or cable, is called tension, **T**. When a rope supports the weight of an object that is at rest, the tension in the rope is equal to the weight of the object:

Equation:

$$T = mg.$$

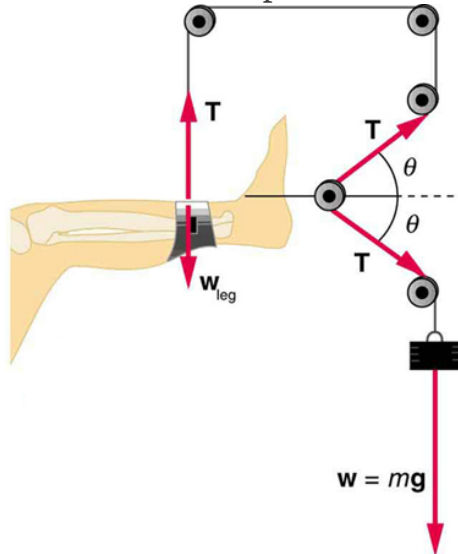
- In any inertial frame of reference (one that is not accelerated or rotated), Newton's laws have the simple forms given in this chapter and all forces are real forces having a physical origin.

Conceptual Questions

Exercise:

Problem:

If a leg is suspended by a traction setup as shown in [\[link\]](#), what is the tension in the rope?



A leg is suspended by a traction system in which wires are used to transmit forces. Frictionless pulleys change the direction of the force T without changing its magnitude.

Exercise:**Problem:**

In a traction setup for a broken bone, with pulleys and rope available, how might we be able to increase the force along the tibia using the same weight? (See [\[link\]](#).) (Note that the tibia is the shin bone shown in this image.)

Problem Exercises

Exercise:

Problem:

Two teams of nine members each engage in a tug of war. Each of the first team's members has an average mass of 68 kg and exerts an average force of 1350 N horizontally. Each of the second team's members has an average mass of 73 kg and exerts an average force of 1365 N horizontally. (a) What is magnitude of the acceleration of the two teams? (b) What is the tension in the section of rope between the teams?

Solution:

- a. 0.11 m/s^2
- b. $1.2 \times 10^4 \text{ N}$

Exercise:

Problem:

What force does a trampoline have to apply to a 45.0-kg gymnast to accelerate her straight up at 7.50 m/s^2 ? Note that the answer is independent of the velocity of the gymnast—she can be moving either up or down, or be stationary.

Exercise:

Problem:

(a) Calculate the tension in a vertical strand of spider web if a spider of mass $8.00 \times 10^{-5} \text{ kg}$ hangs motionless on it. (b) Calculate the tension in a horizontal strand of spider web if the same spider sits motionless in the middle of it much like the tightrope walker in [\[link\]](#). The strand sags at an angle of 12° below the horizontal. Compare this with the tension in the vertical strand (find their ratio).

Solution:

(a) $7.84 \times 10^{-4} \text{ N}$

(b) $1.89 \times 10^{-3} \text{ N}$. This is 2.41 times the tension in the vertical strand.

Exercise:**Problem:**

Suppose a 60.0-kg gymnast climbs a rope. (a) What is the tension in the rope if he climbs at a constant speed? (b) What is the tension in the rope if he accelerates upward at a rate of 1.50 m/s^2 ?

Exercise:**Problem:**

Show that, as stated in the text, a force \mathbf{F}_\perp exerted on a flexible medium at its center and perpendicular to its length (such as on the tightrope wire in [\[link\]](#)) gives rise to a tension of magnitude

$$T = \frac{F_\perp}{2 \sin(\theta)}.$$

Solution:

Newton's second law applied in vertical direction gives

Equation:

$$F_y = F - 2T \sin \theta = 0$$

Equation:

$$F = 2T \sin \theta$$

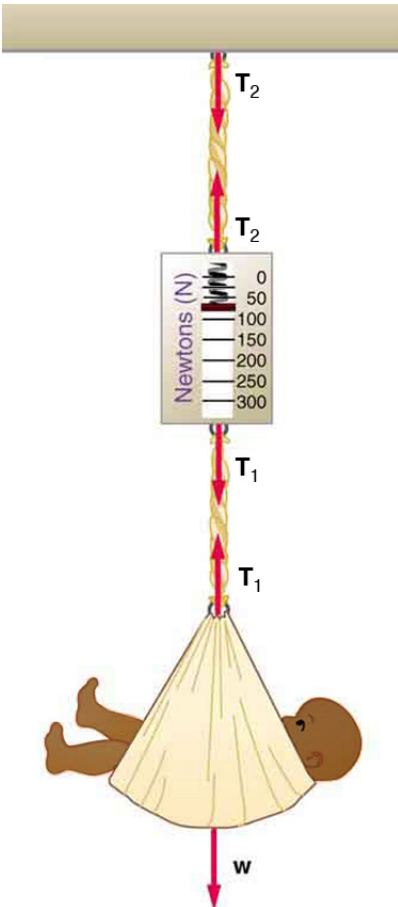
Equation:

$$T = \frac{F}{2 \sin \theta}.$$

Exercise:

Problem:

Consider the baby being weighed in [\[link\]](#). (a) What is the mass of the child and basket if a scale reading of 55 N is observed? (b) What is the tension T_1 in the cord attaching the baby to the scale? (c) What is the tension T_2 in the cord attaching the scale to the ceiling, if the scale has a mass of 0.500 kg? (d) Draw a sketch of the situation indicating the system of interest used to solve each part. The masses of the cords are negligible.



A baby is weighed
using a spring
scale.

Glossary

inertial frame of reference

a coordinate system that is not accelerating; all forces acting in an inertial frame of reference are real forces, as opposed to fictitious forces that are observed due to an accelerating frame of reference

normal force

the force that a surface applies to an object to support the weight of the object; acts perpendicular to the surface on which the object rests

tension

the pulling force that acts along a medium, especially a stretched flexible connector, such as a rope or cable; when a rope supports the weight of an object, the force on the object due to the rope is called a tension force

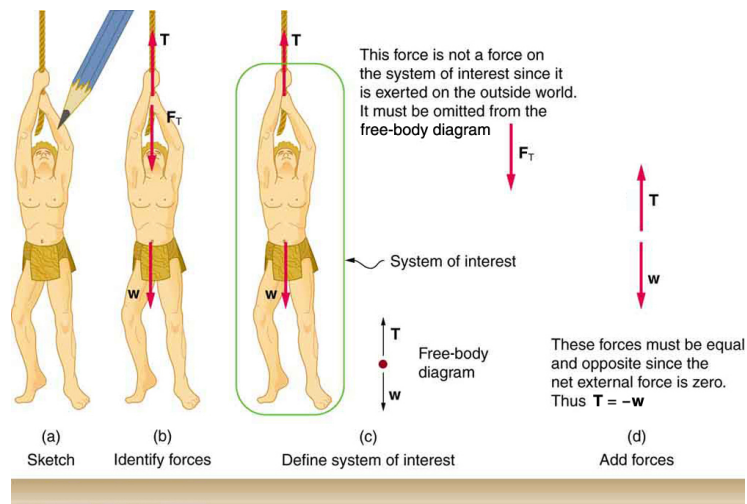
Problem-Solving Strategies

- Understand and apply a problem-solving procedure to solve problems using Newton's laws of motion.

Success in problem solving is obviously necessary to understand and apply physical principles, not to mention the more immediate need of passing exams. The basics of problem solving, presented earlier in this text, are followed here, but specific strategies useful in applying Newton's laws of motion are emphasized. These techniques also reinforce concepts that are useful in many other areas of physics. Many problem-solving strategies are stated outright in the worked examples, and so the following techniques should reinforce skills you have already begun to develop.

Problem-Solving Strategy for Newton's Laws of Motion

Step 1. As usual, it is first necessary to identify the physical principles involved. *Once it is determined that Newton's laws of motion are involved (if the problem involves forces), it is particularly important to draw a careful sketch of the situation.* Such a sketch is shown in [\[link\]\(a\)](#). Then, as in [\[link\]\(b\)](#), use arrows to represent all forces, label them carefully, and make their lengths and directions correspond to the forces they represent (whenever sufficient information exists).



(a) A sketch of Tarzan hanging from a vine. (b) Arrows are used to represent all forces. \mathbf{T} is the tension in the vine above Tarzan, \mathbf{F}_T is the force he exerts on the vine, and \mathbf{w} is his weight. All other forces, such as the nudge of a breeze, are assumed negligible. (c) Suppose we are given the ape man's mass and asked to find the tension in the vine. We then define the system of interest as shown and draw a free-body diagram. \mathbf{F}_T is no longer shown, because it is not a force acting on the system of interest; rather, \mathbf{F}_T acts on the outside world. (d) Showing only the arrows, the head-to-tail method of addition is used. It is apparent that $\mathbf{T} = -\mathbf{w}$, if Tarzan is stationary.

Step 2. Identify what needs to be determined and what is known or can be inferred from the problem as stated. That is, make a list of knowns and unknowns. *Then carefully determine the system of interest.* This decision is a crucial step, since Newton's second law involves only external forces. Once the system of interest has been identified, it becomes possible to determine which forces are external and which are internal, a necessary step to

employ Newton's second law. (See [\[link\]\(c\)](#).) Newton's third law may be used to identify whether forces are exerted between components of a system (internal) or between the system and something outside (external). As illustrated earlier in this chapter, the system of interest depends on what question we need to answer. This choice becomes easier with practice, eventually developing into an almost unconscious process. Skill in clearly defining systems will be beneficial in later chapters as well.

A diagram showing the system of interest and all of the external forces is called a **free-body diagram**. Only forces are shown on free-body diagrams, not acceleration or velocity. We have drawn several of these in worked examples. [\[link\]\(c\)](#) shows a free-body diagram for the system of interest. Note that no internal forces are shown in a free-body diagram.

Step 3. Once a free-body diagram is drawn, *Newton's second law can be applied to solve the problem*. This is done in [\[link\]\(d\)](#) for a particular situation. In general, once external forces are clearly identified in free-body diagrams, it should be a straightforward task to put them into equation form and solve for the unknown, as done in all previous examples. If the problem is one-dimensional—that is, if all forces are parallel—then they add like scalars. If the problem is two-dimensional, then it must be broken down into a pair of one-dimensional problems. This is done by projecting the force vectors onto a set of axes chosen for convenience. As seen in previous examples, the choice of axes can simplify the problem. For example, when an incline is involved, a set of axes with one axis parallel to the incline and one perpendicular to it is most convenient. It is almost always convenient to make one axis parallel to the direction of motion, if this is known.

Note:

Applying Newton's Second Law

Before you write net force equations, it is critical to determine whether the system is accelerating in a particular direction. If the acceleration is zero in a particular direction, then the net force is zero in that direction. Similarly, if the acceleration is nonzero in a particular direction, then the net force is described by the equation: $F_{\text{net}} = ma$. For example, if the system is accelerating in the horizontal direction, but it is not accelerating in the vertical direction, then you will have the following conclusions:

Equation:

$$F_{\text{net } x} = ma,$$

Equation:

$$F_{\text{net } y} = 0.$$

You will need this information in order to determine unknown forces acting in a system.

Step 4. As always, *check the solution to see whether it is reasonable*. In some cases, this is obvious. For example, it is reasonable to find that friction causes an object to slide down an incline more slowly than when no friction exists. In practice, intuition develops gradually through problem solving, and with experience it becomes progressively easier to judge whether an answer is reasonable. Another way to check your solution is to check the units. If you are solving for force and end up with units of m/s, then you have made a mistake.

Section Summary

- To solve problems involving Newton's laws of motion, follow the procedure described:
 1. Draw a sketch of the problem.
 2. Identify known and unknown quantities, and identify the system of interest. Draw a free-body diagram, which is a sketch showing all of the forces acting on an object. The object is represented by a dot, and the forces are represented by vectors extending in different directions from the dot. If vectors act in

directions that are not horizontal or vertical, resolve the vectors into horizontal and vertical components and draw them on the free-body diagram.

3. Write Newton's second law in the horizontal and vertical directions and add the forces acting on the object. If the object does not accelerate in a particular direction (for example, the x -direction) then $F_{\text{net } x} = 0$. If the object does accelerate in that direction, $F_{\text{net } x} = ma$.
4. Check your answer. Is the answer reasonable? Are the units correct?

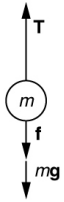
Problem Exercises

Exercise:

Problem:

A 5.00×10^5 -kg rocket is accelerating straight up. Its engines produce 1.250×10^7 N of thrust, and air resistance is 4.50×10^6 N. What is the rocket's acceleration? Explicitly show how you follow the steps in the Problem-Solving Strategy for Newton's laws of motion.

Solution:



Using the free-body diagram:

$$F_{\text{net}} = T - f - mg = ma,$$

so that

$$a = \frac{T - f - mg}{m} = \frac{1.250 \times 10^7 \text{ N} - 4.50 \times 10^6 \text{ N} - (5.00 \times 10^5 \text{ kg})(9.80 \text{ m/s}^2)}{5.00 \times 10^5 \text{ kg}} = 6.20 \text{ m/s}^2.$$

Exercise:

Problem:

The wheels of a midsize car exert a force of 2100 N backward on the road to accelerate the car in the forward direction. If the force of friction including air resistance is 250 N and the acceleration of the car is 1.80 m/s^2 , what is the mass of the car plus its occupants? Explicitly show how you follow the steps in the Problem-Solving Strategy for Newton's laws of motion. For this situation, draw a free-body diagram and write the net force equation.

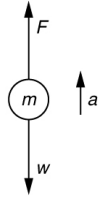
Exercise:

Problem:

Calculate the force a 70.0-kg high jumper must exert on the ground to produce an upward acceleration 4.00 times the acceleration due to gravity. Explicitly show how you follow the steps in the Problem-Solving Strategy for Newton's laws of motion.

Solution:

Use Newton's laws of motion.



Given : $a = 4.00g = (4.00)(9.80 \text{ m/s}^2) = 39.2 \text{ m/s}^2$; $m = 70.0 \text{ kg}$,

Find: F .

$$\sum F = +F - w = ma, \text{ so } F = ma + w = ma + mg = m(a + g).$$

that

$$F = (70.0 \text{ kg})[(39.2 \text{ m/s}^2) + (9.80 \text{ m/s}^2)] = 3.43 \times 10^3 \text{ N}.$$

The force exerted by the high-jumper is actually down on the ground, but F is up from the ground and makes him jump.

This result is reasonable, since it is quite possible for a person to exert a force of the magnitude of 10^3 N .

Exercise:

Problem:

When landing after a spectacular somersault, a 40.0-kg gymnast decelerates by pushing straight down on the mat. Calculate the force she must exert if her deceleration is 7.00 times the acceleration due to gravity.

Explicitly show how you follow the steps in the Problem-Solving Strategy for Newton's laws of motion.

Exercise:

Problem:

A freight train consists of two 8.00×10^4 -kg engines and 45 cars with average masses of $5.50 \times 10^4 \text{ kg}$. (a) What force must each engine exert backward on the track to accelerate the train at a rate of $5.00 \times 10^{-2} \text{ m/s}^2$ if the force of friction is $7.50 \times 10^5 \text{ N}$, assuming the engines exert identical forces? This is not a large frictional force for such a massive system. Rolling friction for trains is small, and consequently trains are very energy-efficient transportation systems. (b) What is the force in the coupling between the 37th and 38th cars (this is the force each exerts on the other), assuming all cars have the same mass and that friction is evenly distributed among all of the cars and engines?

Solution:

(a) $4.41 \times 10^5 \text{ N}$

(b) $1.50 \times 10^5 \text{ N}$

Exercise:

Problem:

Commercial airplanes are sometimes pushed out of the passenger loading area by a tractor. (a) An 1800-kg tractor exerts a force of $1.75 \times 10^4 \text{ N}$ backward on the pavement, and the system experiences forces resisting motion that total 2400 N. If the acceleration is 0.150 m/s^2 , what is the mass of the airplane? (b) Calculate the force exerted by the tractor on the airplane, assuming 2200 N of the friction is experienced by the airplane. (c) Draw two sketches showing the systems of interest used to solve each part, including the free-body diagrams for each.

Exercise:

Problem:

A 1100-kg car pulls a boat on a trailer. (a) What total force resists the motion of the car, boat, and trailer, if the car exerts a 1900-N force on the road and produces an acceleration of 0.550 m/s^2 ? The mass of the boat plus trailer is 700 kg. (b) What is the force in the hitch between the car and the trailer if 80% of the resisting forces are experienced by the boat and trailer?

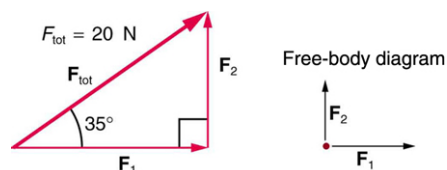
Solution:

(a) 910 N

(b) $1.11 \times 10^3 \text{ N}$

Exercise:**Problem:**

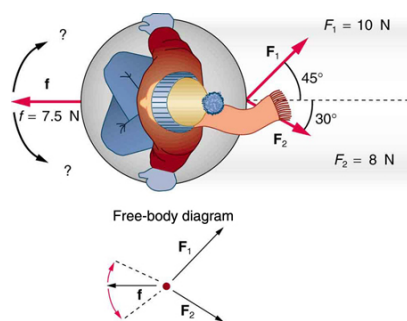
(a) Find the magnitudes of the forces \mathbf{F}_1 and \mathbf{F}_2 that add to give the total force \mathbf{F}_{tot} shown in [\[link\]](#). This may be done either graphically or by using trigonometry. (b) Show graphically that the same total force is obtained independent of the order of addition of \mathbf{F}_1 and \mathbf{F}_2 . (c) Find the direction and magnitude of some other pair of vectors that add to give \mathbf{F}_{tot} . Draw these to scale on the same drawing used in part (b) or a similar picture.

**Exercise:****Problem:**

Two children pull a third child on a snow saucer sled exerting forces \mathbf{F}_1 and \mathbf{F}_2 as shown from above in [\[link\]](#). Find the acceleration of the 49.00-kg sled and child system. Note that the direction of the frictional force is unspecified; it will be in the opposite direction of the sum of \mathbf{F}_1 and \mathbf{F}_2 .

Solution:

$a = 0.139 \text{ m/s}$, $\theta = 12.4^\circ$ north of east



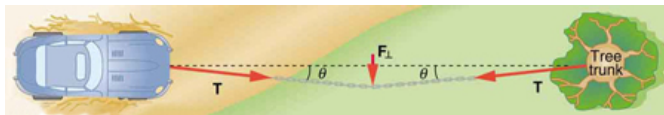
An overhead view of the horizontal forces acting on a

child's snow saucer sled.

Exercise:

Problem:

Suppose your car was mired deeply in the mud and you wanted to use the method illustrated in [\[link\]](#) to pull it out. (a) What force would you have to exert perpendicular to the center of the rope to produce a force of 12,000 N on the car if the angle is 2.00° ? In this part, explicitly show how you follow the steps in the Problem-Solving Strategy for Newton's laws of motion. (b) Real ropes stretch under such forces. What force would be exerted on the car if the angle increases to 7.00° and you still apply the force found in part (a) to its center?



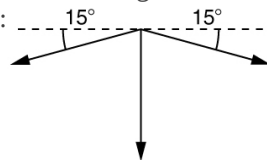
Exercise:

Problem:

What force is exerted on the tooth in [\[link\]](#) if the tension in the wire is 25.0 N? Note that the force applied to the tooth is smaller than the tension in the wire, but this is necessitated by practical considerations of how force can be applied in the mouth. Explicitly show how you follow steps in the Problem-Solving Strategy for Newton's laws of motion.

Solution:

Use Newton's laws since we are looking for forces.
Draw a free-body diagram:

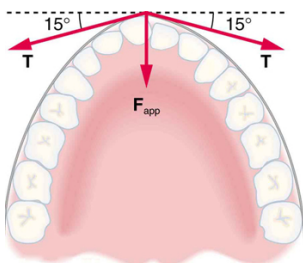


The tension is given as $T = 25.0 \text{ N}$. Find F_{app} . Using Newton's laws gives:

$\Sigma F_y = 0$, so that y-components of the two tensions:

$F_{\text{app}} = 2 T \sin \theta = 2(25.0 \text{ N}) \sin(15^\circ) =$

This seems reasonable, since the applied tensions should be greater than the force applied to the tooth.



Braces are used to apply forces to teeth to realign them. Shown in this figure are the tensions applied by the wire to the protruding tooth. The total force applied to the tooth by the wire, \mathbf{F}_{app} , points straight toward the back of the mouth.

Exercise:

Problem:

[\[link\]](#) shows Superhero and Trusty Sidekick hanging motionless from a rope. Superhero's mass is 90.0 kg, while Trusty Sidekick's is 55.0 kg, and the mass of the rope is negligible. (a) Draw a free-body diagram of the situation showing all forces acting on Superhero, Trusty Sidekick, and the rope. (b) Find the tension in the rope above Superhero. (c) Find the tension in the rope between Superhero and Trusty Sidekick. Indicate on your free-body diagram the system of interest used to solve each part.



Superhero and Trusty Sidekick hang motionless on a rope as they try to figure out what to do next. Will the tension be the same everywhere in the rope?

Exercise:**Problem:**

A nurse pushes a cart by exerting a force on the handle at a downward angle 35.0° below the horizontal. The loaded cart has a mass of 28.0 kg, and the force of friction is 60.0 N. (a) Draw a free-body diagram for the system of interest. (b) What force must the nurse exert to move at a constant velocity?

Exercise:**Problem:**

Construct Your Own Problem Consider the tension in an elevator cable during the time the elevator starts from rest and accelerates its load upward to some cruising velocity. Taking the elevator and its load to be the system of interest, draw a free-body diagram. Then calculate the tension in the cable. Among the things to consider are the mass of the elevator and its load, the final velocity, and the time taken to reach that velocity.

Exercise:**Problem:**

Construct Your Own Problem Consider two people pushing a toboggan with four children on it up a snow-covered slope. Construct a problem in which you calculate the acceleration of the toboggan and its load. Include a free-body diagram of the appropriate system of interest as the basis for your analysis. Show vector forces and their components and explain the choice of coordinates. Among the things to be considered are the forces exerted by those pushing, the angle of the slope, and the masses of the toboggan and children.

Exercise:**Problem:**

Unreasonable Results (a) Repeat [\[link\]](#), but assume an acceleration of 1.20 m/s^2 is produced. (b) What is unreasonable about the result? (c) Which premise is unreasonable, and why is it unreasonable?

Exercise:**Problem:**

Unreasonable Results (a) What is the initial acceleration of a rocket that has a mass of $1.50 \times 10^6 \text{ kg}$ at takeoff, the engines of which produce a thrust of $2.00 \times 10^6 \text{ N}$? Do not neglect gravity. (b) What is unreasonable about the result? (This result has been unintentionally achieved by several real rockets.) (c) Which premise is unreasonable, or which premises are inconsistent? (You may find it useful to compare this problem to the rocket problem earlier in this section.)

Further Applications of Newton's Laws of Motion

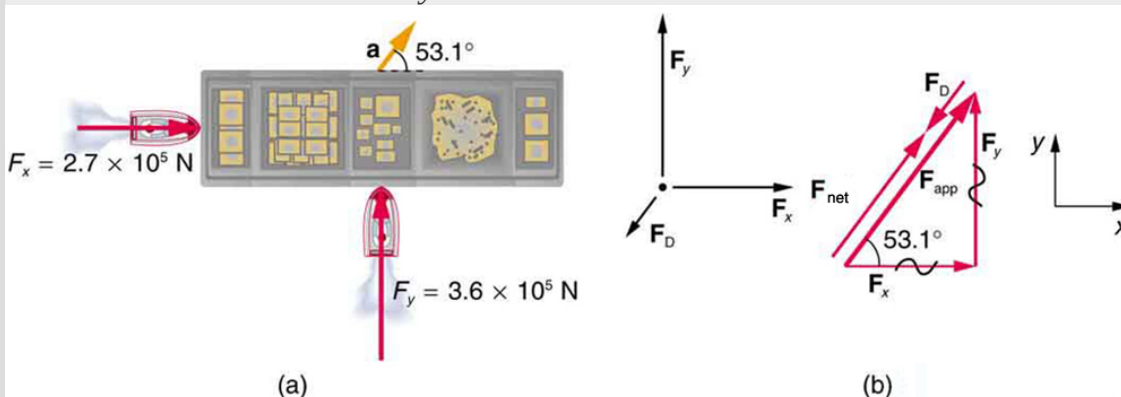
- Apply problem-solving techniques to solve for quantities in more complex systems of forces.
- Integrate concepts from kinematics to solve problems using Newton's laws of motion.

There are many interesting applications of Newton's laws of motion, a few more of which are presented in this section. These serve also to illustrate some further subtleties of physics and to help build problem-solving skills.

Example:

Drag Force on a Barge

Suppose two tugboats push on a barge at different angles, as shown in [\[link\]](#). The first tugboat exerts a force of $2.7 \times 10^5 \text{ N}$ in the x -direction, and the second tugboat exerts a force of $3.6 \times 10^5 \text{ N}$ in the y -direction.



(a) A view from above of two tugboats pushing on a barge. (b) The free-body diagram for the ship contains only forces acting in the plane of the water. It omits the two vertical forces—the weight of the barge and the buoyant force of the water supporting it cancel and are not shown. Since the applied forces are perpendicular, the x - and y -axes are in the same direction as \mathbf{F}_x and \mathbf{F}_y . The problem quickly becomes a one-dimensional problem along the direction of \mathbf{F}_{app} , since friction is in the direction opposite to \mathbf{F}_{app} .

If the mass of the barge is $5.0 \times 10^6 \text{ kg}$ and its acceleration is observed to be $7.5 \times 10^{-2} \text{ m/s}^2$ in the direction shown, what is the drag force of the water on the

barge resisting the motion? (Note: drag force is a frictional force exerted by fluids, such as air or water. The drag force opposes the motion of the object.)

Strategy

The directions and magnitudes of acceleration and the applied forces are given in [\[link\]\(a\)](#). We will define the total force of the tugboats on the barge as \mathbf{F}_{app} so that:

Equation:

$$\mathbf{F}_{\text{app}} = \mathbf{F}_x + \mathbf{F}_y$$

Since the barge is flat bottomed, the drag of the water \mathbf{F}_D will be in the direction opposite to \mathbf{F}_{app} , as shown in the free-body diagram in [\[link\]\(b\)](#). The system of interest here is the barge, since the forces on *it* are given as well as its acceleration. Our strategy is to find the magnitude and direction of the net applied force \mathbf{F}_{app} , and then apply Newton's second law to solve for the drag force \mathbf{F}_D .

Solution

Since \mathbf{F}_x and \mathbf{F}_y are perpendicular, the magnitude and direction of \mathbf{F}_{app} are easily found. First, the resultant magnitude is given by the Pythagorean theorem:

Equation:

$$\begin{aligned} F_{\text{app}} &= \sqrt{F_x^2 + F_y^2} \\ F_{\text{app}} &= \sqrt{(2.7 \times 10^5 \text{ N})^2 + (3.6 \times 10^5 \text{ N})^2} = 4.5 \times 10^5 \text{ N}. \end{aligned}$$

The angle is given by

Equation:

$$\begin{aligned} \theta &= \tan^{-1} \left(\frac{F_y}{F_x} \right) \\ \theta &= \tan^{-1} \left(\frac{3.6 \times 10^5 \text{ N}}{2.7 \times 10^5 \text{ N}} \right) = 53^\circ, \end{aligned}$$

which we know, because of Newton's first law, is the same direction as the acceleration. \mathbf{F}_D is in the opposite direction of \mathbf{F}_{app} , since it acts to slow down the acceleration. Therefore, the net external force is in the same direction as \mathbf{F}_{app} , but its magnitude is slightly less than \mathbf{F}_{app} . The problem is now one-dimensional. From [\[link\]\(b\)](#), we can see that

Equation:

$$F_{\text{net}} = F_{\text{app}} - F_D.$$

But Newton's second law states that

Equation:

$$F_{\text{net}} = ma.$$

Thus,

Equation:

$$F_{\text{app}} - F_{\text{D}} = ma.$$

This can be solved for the magnitude of the drag force of the water F_{D} in terms of known quantities:

Equation:

$$F_{\text{D}} = F_{\text{app}} - ma.$$

Substituting known values gives

Equation:

$$F_{\text{D}} = (4.5 \times 10^5 \text{ N}) - (5.0 \times 10^6 \text{ kg})(7.5 \times 10^{-2} \text{ m/s}^2) = 7.5 \times 10^4 \text{ N}.$$

The direction of \mathbf{F}_{D} has already been determined to be in the direction opposite to \mathbf{F}_{app} , or at an angle of 53° south of west.

Discussion

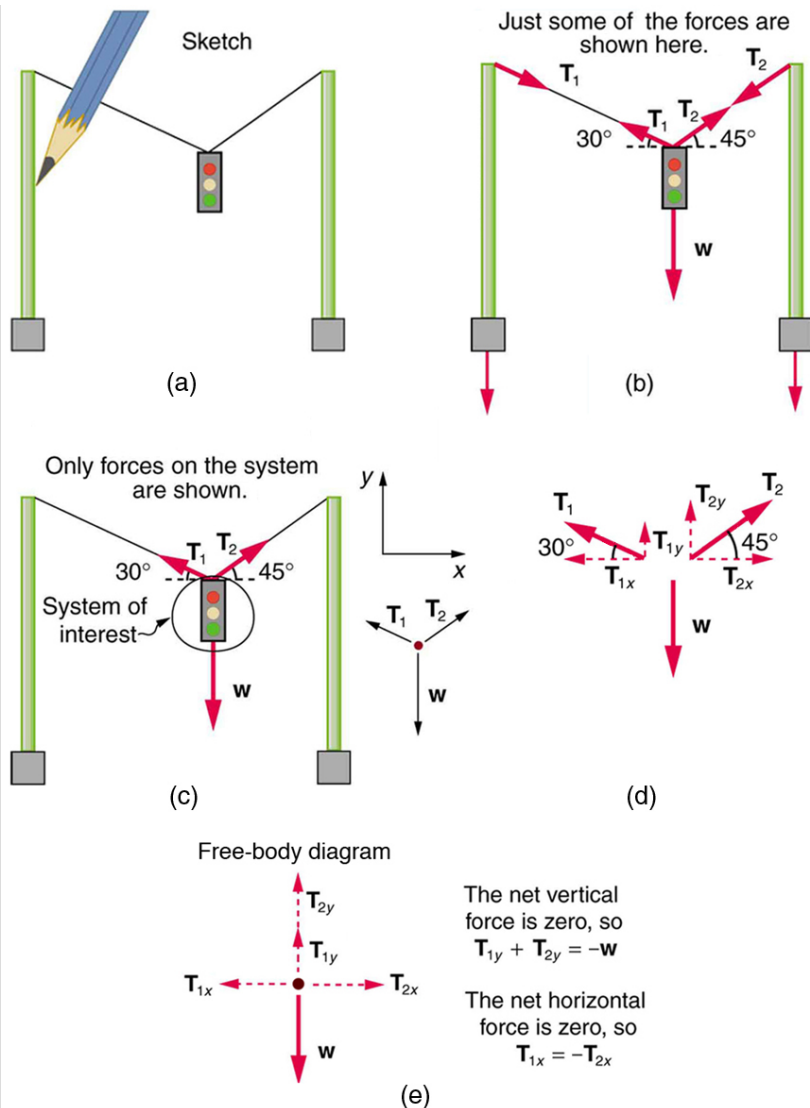
The numbers used in this example are reasonable for a moderately large barge. It is certainly difficult to obtain larger accelerations with tugboats, and small speeds are desirable to avoid running the barge into the docks. Drag is relatively small for a well-designed hull at low speeds, consistent with the answer to this example, where F_{D} is less than 1/600th of the weight of the ship.

In the earlier example of a tightrope walker we noted that the tensions in wires supporting a mass were equal only because the angles on either side were equal. Consider the following example, where the angles are not equal; slightly more trigonometry is involved.

Example:

Different Tensions at Different Angles

Consider the traffic light (mass 15.0 kg) suspended from two wires as shown in [\[link\]](#). Find the tension in each wire, neglecting the masses of the wires.



A traffic light is suspended from two wires. (b) Some of the forces involved. (c) Only forces acting on the system are shown here. The free-body diagram for the traffic light is also shown. (d) The forces projected onto vertical (y) and horizontal (x) axes. The horizontal components of the tensions must cancel, and the sum of the vertical components of the tensions must equal the weight of the traffic light. (e) The free-body diagram shows the vertical and horizontal forces acting on the traffic light.

Strategy

The system of interest is the traffic light, and its free-body diagram is shown in [\[link\]](#) (c). The three forces involved are not parallel, and so they must be projected onto a coordinate system. The most convenient coordinate system has one axis vertical and one horizontal, and the vector projections on it are shown in part (d) of the figure. There are two unknowns in this problem (T_1 and T_2), so two equations are needed to find them. These two equations come from applying Newton's second law along the vertical and horizontal axes, noting that the net external force is zero along each axis because acceleration is zero.

Solution

First consider the horizontal or x -axis:

Equation:

$$F_{\text{net}x} = T_{2x} - T_{1x} = 0.$$

Thus, as you might expect,

Equation:

$$T_{1x} = T_{2x}.$$

This gives us the following relationship between T_1 and T_2 :

Equation:

$$T_1 \cos(30^\circ) = T_2 \cos(45^\circ).$$

Thus,

Equation:

$$T_2 = (1.225)T_1.$$

Note that T_1 and T_2 are not equal in this case, because the angles on either side are not equal. It is reasonable that T_2 ends up being greater than T_1 , because it is exerted more vertically than T_1 .

Now consider the force components along the vertical or y -axis:

Equation:

$$F_{\text{net}y} = T_{1y} + T_{2y} - w = 0.$$

This implies

Equation:

$$T_{1y} + T_{2y} = w.$$

Substituting the expressions for the vertical components gives

Equation:

$$T_1 \sin (30^\circ) + T_2 \sin (45^\circ) = w.$$

There are two unknowns in this equation, but substituting the expression for T_2 in terms of T_1 reduces this to one equation with one unknown:

Equation:

$$T_1(0.500) + (1.225T_1)(0.707) = w = mg,$$

which yields

Equation:

$$(1.366)T_1 = (15.0 \text{ kg})(9.80 \text{ m/s}^2).$$

Solving this last equation gives the magnitude of T_1 to be

Equation:

$$T_1 = 108 \text{ N}.$$

Finally, the magnitude of T_2 is determined using the relationship between them, $T_2 = 1.225 T_1$, found above. Thus we obtain

Equation:

$$T_2 = 132 \text{ N}.$$

Discussion

Both tensions would be larger if both wires were more horizontal, and they will be equal if and only if the angles on either side are the same (as they were in the earlier example of a tightrope walker).

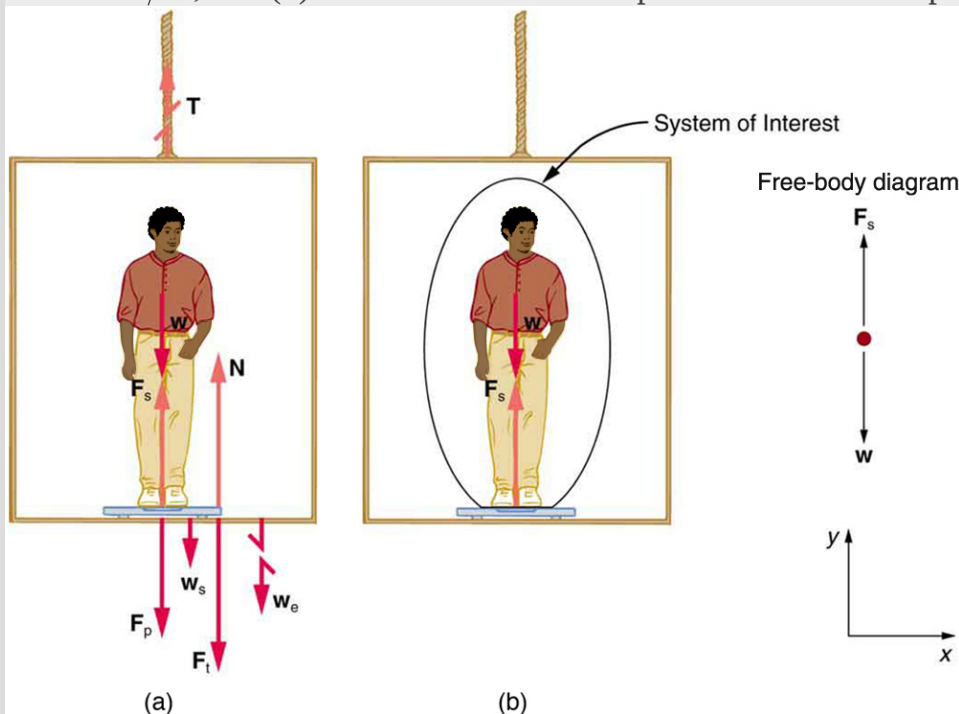
The bathroom scale is an excellent example of a normal force acting on a body. It provides a quantitative reading of how much it must push upward to support the weight of an object. But can you predict what you would see on the dial of a bathroom scale if you stood on it during an elevator ride? Will you see a value greater than your weight when the elevator starts up? What about when the elevator moves upward at a constant speed: will the scale still read more than your weight at rest? Consider the following example.

Example:

What Does the Bathroom Scale Read in an Elevator?

[\[link\]](#) shows a 75.0-kg man (weight of about 165 lb) standing on a bathroom scale in an elevator. Calculate the scale reading: (a) if the elevator accelerates upward at a rate

of 1.20 m/s^2 , and (b) if the elevator moves upward at a constant speed of 1 m/s .



(a) The various forces acting when a person stands on a bathroom scale in an elevator. The arrows are approximately correct for when the elevator is accelerating upward—broken arrows represent forces too large to be drawn to scale. \mathbf{T} is the tension in the supporting cable, \mathbf{w} is the weight of the person, \mathbf{w}_s is the weight of the scale, \mathbf{w}_e is the weight of the elevator, \mathbf{F}_s is the force of the scale on the person, \mathbf{F}_p is the force of the person on the scale, \mathbf{F}_t is the force of the scale on the floor of the elevator, and \mathbf{N} is the force of the floor upward on the scale. (b) The free-body diagram shows only the external forces acting on the designated system of interest—the person.

Strategy

If the scale is accurate, its reading will equal F_p , the magnitude of the force the person exerts downward on it. [\[link\]](#)(a) shows the numerous forces acting on the elevator, scale, and person. It makes this one-dimensional problem look much more formidable than if the person is chosen to be the system of interest and a free-body diagram is drawn as in [\[link\]](#)(b). Analysis of the free-body diagram using Newton's laws can produce answers to both parts (a) and (b) of this example, as well as some other questions that might arise. The only forces acting on the person are his weight \mathbf{w} and the upward force of the scale \mathbf{F}_s . According to Newton's third law \mathbf{F}_p and \mathbf{F}_s are

equal in magnitude and opposite in direction, so that we need to find F_s in order to find what the scale reads. We can do this, as usual, by applying Newton's second law,
Equation:

$$F_{\text{net}} = ma.$$

From the free-body diagram we see that $F_{\text{net}} = F_s - w$, so that
Equation:

$$F_s - w = ma.$$

Solving for F_s gives an equation with only one unknown:
Equation:

$$F_s = ma + w,$$

or, because $w = mg$, simply
Equation:

$$F_s = ma + mg.$$

No assumptions were made about the acceleration, and so this solution should be valid for a variety of accelerations in addition to the ones in this exercise.

Solution for (a)

In this part of the problem, $a = 1.20 \text{ m/s}^2$, so that

Equation:

$$F_s = (75.0 \text{ kg})(1.20 \text{ m/s}^2) + (75.0 \text{ kg})(9.80 \text{ m/s}^2),$$

yielding

Equation:

$$F_s = 825 \text{ N}.$$

Discussion for (a)

This is about 185 lb. What would the scale have read if he were stationary? Since his acceleration would be zero, the force of the scale would be equal to his weight:

Equation:

$$\begin{aligned} F_{\text{net}} &= ma = 0 = F_s - w \\ F_s &= w = mg \\ F_s &= (75.0 \text{ kg})(9.80 \text{ m/s}^2) \\ F_s &= 735 \text{ N}. \end{aligned}$$

So, the scale reading in the elevator is greater than his 735-N (165 lb) weight. This means that the scale is pushing up on the person with a force greater than his weight, as it must in order to accelerate him upward. Clearly, the greater the acceleration of the elevator, the greater the scale reading, consistent with what you feel in rapidly accelerating versus slowly accelerating elevators.

Solution for (b)

Now, what happens when the elevator reaches a constant upward velocity? Will the scale still read more than his weight? For any constant velocity—up, down, or stationary—acceleration is zero because $a = \frac{\Delta v}{\Delta t}$, and $\Delta v = 0$.

Thus,

Equation:

$$F_s = ma + mg = 0 + mg.$$

Now

Equation:

$$F_s = (75.0 \text{ kg})(9.80 \text{ m/s}^2),$$

which gives

Equation:

$$F_s = 735 \text{ N}.$$

Discussion for (b)

The scale reading is 735 N, which equals the person's weight. This will be the case whenever the elevator has a constant velocity—moving up, moving down, or stationary.

The solution to the previous example also applies to an elevator accelerating downward, as mentioned. When an elevator accelerates downward, a is negative, and the scale reading is *less* than the weight of the person, until a constant downward velocity is reached, at which time the scale reading again becomes equal to the person's weight. If the elevator is in free-fall and accelerating downward at g , then the scale reading will be zero and the person will *appear* to be weightless.

Integrating Concepts: Newton's Laws of Motion and Kinematics

Physics is most interesting and most powerful when applied to general situations that involve more than a narrow set of physical principles. Newton's laws of motion can also be integrated with other concepts that have been discussed previously in this text to

solve problems of motion. For example, forces produce accelerations, a topic of kinematics, and hence the relevance of earlier chapters. When approaching problems that involve various types of forces, acceleration, velocity, and/or position, use the following steps to approach the problem:

Problem-Solving Strategy

- Step 1. *Identify which physical principles are involved.* Listing the givens and the quantities to be calculated will allow you to identify the principles involved.
- Step 2. *Solve the problem using strategies outlined in the text.* If these are available for the specific topic, you should refer to them. You should also refer to the sections of the text that deal with a particular topic. The following worked example illustrates how these strategies are applied to an integrated concept problem.

Example:

What Force Must a Soccer Player Exert to Reach Top Speed?

A soccer player starts from rest and accelerates forward, reaching a velocity of 8.00 m/s in 2.50 s. (a) What was his average acceleration? (b) What average force did he exert backward on the ground to achieve this acceleration? The player’s mass is 70.0 kg, and air resistance is negligible.

Strategy

To solve an integrated concept problem, first identify the physical principles involved and identify the chapters in which they are found. Part (a) of this example considers acceleration along a straight line. This is a topic of kinematics. Part (b) deals with force, a dynamics topic found in this chapter.

The following solutions to each part of the example illustrate how the specific

problem-solving strategies are applied. These involve identifying knowns and unknowns, checking to see if the answer is reasonable, and so forth.

Solution for (a)

We are given the initial and final velocities (zero and 8.00 m/s forward); thus, the change in velocity is $\Delta v = 8.00 \text{ m/s}$. We are given the elapsed time, and so $\Delta t = 2.50 \text{ s}$. The unknown is acceleration, which can be found from its definition:

Equation:

$$a = \frac{\Delta v}{\Delta t}.$$

Substituting the known values yields

Equation:

$$\begin{aligned} a &= \frac{8.00 \text{ m/s}}{2.50 \text{ s}} \\ &= 3.20 \text{ m/s}^2. \end{aligned}$$

Discussion for (a)

This is an attainable acceleration for an athlete in good condition.

Solution for (b)

Here we are asked to find the average force the player exerts backward to achieve this forward acceleration. Neglecting air resistance, this would be equal in magnitude to the net external force on the player, since this force causes his acceleration. Since we now know the player's acceleration and are given his mass, we can use Newton's second law to find the force exerted. That is,

Equation:

$$F_{\text{net}} = ma.$$

Substituting the known values of m and a gives

Equation:

$$\begin{aligned} F_{\text{net}} &= (70.0 \text{ kg})(3.20 \text{ m/s}^2) \\ &= 224 \text{ N}. \end{aligned}$$

Discussion for (b)

This is about 50 pounds, a reasonable average force.

This worked example illustrates how to apply problem-solving strategies to situations that include topics from different chapters. The first step is to identify the physical principles involved in the problem. The second step is to solve for the unknown using familiar problem-solving strategies. These strategies are found throughout the text, and many worked examples show how to use them for single topics. You will find these

techniques for integrated concept problems useful in applications of physics outside of a physics course, such as in your profession, in other science disciplines, and in everyday life. The following problems will build your skills in the broad application of physical principles.

Summary

- Newton's laws of motion can be applied in numerous situations to solve problems of motion.
- Some problems will contain multiple force vectors acting in different directions on an object. Be sure to draw diagrams, resolve all force vectors into horizontal and vertical components, and draw a free-body diagram. Always analyze the direction in which an object accelerates so that you can determine whether $F_{\text{net}} = ma$ or $F_{\text{net}} = 0$.
- The normal force on an object is not always equal in magnitude to the weight of the object. If an object is accelerating, the normal force will be less than or greater than the weight of the object. Also, if the object is on an inclined plane, the normal force will always be less than the full weight of the object.
- Some problems will contain various physical quantities, such as forces, acceleration, velocity, or position. You can apply concepts from kinematics and dynamics in order to solve these problems of motion.

Conceptual Questions

Exercise:

Problem:

To simulate the apparent weightlessness of space orbit, astronauts are trained in the hold of a cargo aircraft that is accelerating downward at g . Why will they appear to be weightless, as measured by standing on a bathroom scale, in this accelerated frame of reference? Is there any difference between their apparent weightlessness in orbit and in the aircraft?

Exercise:

Problem:

A cartoon shows the toupee coming off the head of an elevator passenger when the elevator rapidly stops during an upward ride. Can this really happen without the person being tied to the floor of the elevator? Explain your answer.

Problem Exercises

Exercise:

Problem:

A flea jumps by exerting a force of 1.20×10^{-5} N straight down on the ground. A breeze blowing on the flea parallel to the ground exerts a force of 0.500×10^{-6} N on the flea. Find the direction and magnitude of the acceleration of the flea if its mass is 6.00×10^{-7} kg. Do not neglect the gravitational force.

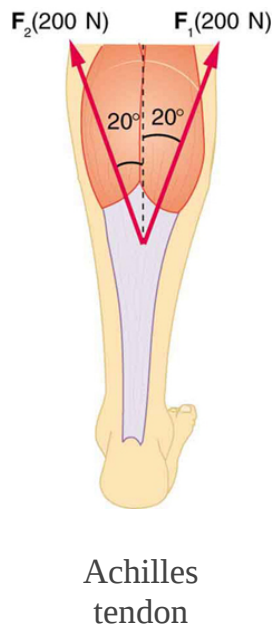
Solution:

10.2 m/s^2 , 4.67° from vertical

Exercise:

Problem:

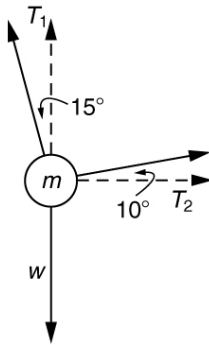
Two muscles in the back of the leg pull upward on the Achilles tendon, as shown in [\[link\]](#). (These muscles are called the medial and lateral heads of the gastrocnemius muscle.) Find the magnitude and direction of the total force on the Achilles tendon. What type of movement could be caused by this force?



Exercise:

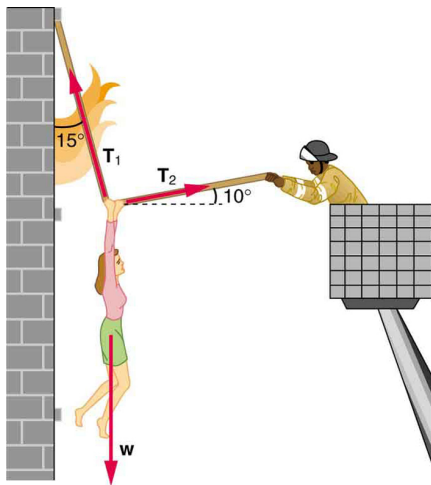
Problem:

A 76.0-kg person is being pulled away from a burning building as shown in [\[link\]](#). Calculate the tension in the two ropes if the person is momentarily motionless. Include a free-body diagram in your solution.

Solution:

$$T_1 = 736 \text{ N}$$

$$T_2 = 194 \text{ N}$$



The force T_2 needed to hold steady the person being rescued from the fire is less than her weight and less than the force T_1 in the other rope, since the more

vertical rope supports a greater part of her weight (a vertical force).

Exercise:

Problem:

Integrated Concepts A 35.0-kg dolphin decelerates from 12.0 to 7.50 m/s in 2.30 s to join another dolphin in play. What average force was exerted to slow him if he was moving horizontally? (The gravitational force is balanced by the buoyant force of the water.)

Exercise:

Problem:

Integrated Concepts When starting a foot race, a 70.0-kg sprinter exerts an average force of 650 N backward on the ground for 0.800 s. (a) What is his final speed? (b) How far does he travel?

Solution:

(a) 7.43 m/s

(b) 2.97 m

Exercise:

Problem:

Integrated Concepts A large rocket has a mass of 2.00×10^6 kg at takeoff, and its engines produce a thrust of 3.50×10^7 N. (a) Find its initial acceleration if it takes off vertically. (b) How long does it take to reach a velocity of 120 km/h straight up, assuming constant mass and thrust? (c) In reality, the mass of a rocket decreases significantly as its fuel is consumed. Describe qualitatively how this affects the acceleration and time for this motion.

Exercise:

Problem:

Integrated Concepts A basketball player jumps straight up for a ball. To do this, he lowers his body 0.300 m and then accelerates through this distance by forcefully straightening his legs. This player leaves the floor with a vertical velocity sufficient to carry him 0.900 m above the floor. (a) Calculate his velocity when he leaves the floor. (b) Calculate his acceleration while he is straightening his legs. He goes from zero to the velocity found in part (a) in a distance of 0.300 m. (c) Calculate the force he exerts on the floor to do this, given that his mass is 110 kg.

Solution:

(a) 4.20 m/s

(b) 29.4 m/s²

(c) 4.31×10^3 N

Exercise:**Problem:**

Integrated Concepts A 2.50-kg fireworks shell is fired straight up from a mortar and reaches a height of 110 m. (a) Neglecting air resistance (a poor assumption, but we will make it for this example), calculate the shell's velocity when it leaves the mortar. (b) The mortar itself is a tube 0.450 m long. Calculate the average acceleration of the shell in the tube as it goes from zero to the velocity found in (a). (c) What is the average force on the shell in the mortar? Express your answer in newtons and as a ratio to the weight of the shell.

Exercise:**Problem:**

Integrated Concepts Repeat [\[link\]](#) for a shell fired at an angle 10.0° from the vertical.

Solution:

(a) 47.1 m/s

(b) 2.47×10^3 m/s²

(c) 6.18×10^3 N . The average force is 252 times the shell's weight.

Exercise:

Problem:

Integrated Concepts An elevator filled with passengers has a mass of 1700 kg. (a) The elevator accelerates upward from rest at a rate of 1.20 m/s^2 for 1.50 s. Calculate the tension in the cable supporting the elevator. (b) The elevator continues upward at constant velocity for 8.50 s. What is the tension in the cable during this time? (c) The elevator decelerates at a rate of 0.600 m/s^2 for 3.00 s. What is the tension in the cable during deceleration? (d) How high has the elevator moved above its original starting point, and what is its final velocity?

Exercise:

Problem:

Unreasonable Results (a) What is the final velocity of a car originally traveling at 50.0 km/h that decelerates at a rate of 0.400 m/s^2 for 50.0 s? (b) What is unreasonable about the result? (c) Which premise is unreasonable, or which premises are inconsistent?

Exercise:

Problem:

Unreasonable Results A 75.0-kg man stands on a bathroom scale in an elevator that accelerates from rest to 30.0 m/s in 2.00 s. (a) Calculate the scale reading in newtons and compare it with his weight. (The scale exerts an upward force on him equal to its reading.) (b) What is unreasonable about the result? (c) Which premise is unreasonable, or which premises are inconsistent?

Extended Topic: The Four Basic Forces—An Introduction

- Understand the four basic forces that underlie the processes in nature.

One of the most remarkable simplifications in physics is that only four distinct forces account for all known phenomena. In fact, nearly all of the forces we experience directly are due to only one basic force, called the electromagnetic force. (The gravitational force is the only force we experience directly that is not electromagnetic.) This is a tremendous simplification of the myriad of *apparently* different forces we can list, only a few of which were discussed in the previous section. As we will see, the basic forces are all thought to act through the exchange of microscopic carrier particles, and the characteristics of the basic forces are determined by the types of particles exchanged. Action at a distance, such as the gravitational force of Earth on the Moon, is explained by the existence of a **force field** rather than by “physical contact.”

The *four basic forces* are the gravitational force, the electromagnetic force, the weak nuclear force, and the strong nuclear force. Their properties are summarized in [\[link\]](#). Since the weak and strong nuclear forces act over an extremely short range, the size of a nucleus or less, we do not experience them directly, although they are crucial to the very structure of matter. These forces determine which nuclei are stable and which decay, and they are the basis of the release of energy in certain nuclear reactions. Nuclear forces determine not only the stability of nuclei, but also the relative abundance of elements in nature. The properties of the nucleus of an atom determine the number of electrons it has and, thus, indirectly determine the chemistry of the atom. More will be said of all of these topics in later chapters.

Note:
Concept Connections: The Four Basic Forces

The four basic forces will be encountered in more detail as you progress through the text. The gravitational force is defined in [Uniform Circular Motion and Gravitation](#), electric force in [Electric Charge and Electric Field](#), magnetic force in [Magnetism](#), and nuclear forces in [Radioactivity and Nuclear Physics](#). On a macroscopic scale, electromagnetism and gravity are the basis for all forces. The nuclear forces are vital to the substructure of matter, but they are not directly experienced on the macroscopic scale.

Force	Approximate Relative Strengths	Range	Attraction/Repulsion	Carrier Particle
Gravitational	10^{-38}	∞	attractive only	Graviton

Force	Approximate Relative Strengths	Range	Attraction/Repulsion	Carrier Particle
Electromagnetic	10^{-2}	∞	attractive and repulsive	Photon
Weak nuclear	10^{-13}	$< 10^{-18}\text{m}$	attractive and repulsive	W^+ , W^- , Z^0
Strong nuclear	1	$< 10^{-15}\text{m}$	attractive and repulsive	gluons

Properties of the Four Basic Forces^[footnote]

The graviton is a proposed particle, though it has not yet been observed by scientists. See the discussion of gravitational waves later in this section. The particles W^+ , W^- , and Z^0 are called vector bosons; these were predicted by theory and first observed in 1983. There are eight types of gluons proposed by scientists, and their existence is indicated by meson exchange in the nuclei of atoms.

The gravitational force is surprisingly weak—it is only because gravity is always attractive that we notice it at all. Our weight is the gravitational force due to the *entire* Earth acting on us. On the very large scale, as in astronomical systems, the gravitational force is the dominant force determining the motions of moons, planets, stars, and galaxies. The gravitational force also affects the nature of space and time. As we shall see later in the study of general relativity, space is curved in the vicinity of very massive bodies, such as the Sun, and time actually slows down near massive bodies.

Electromagnetic forces can be either attractive or repulsive. They are long-range forces, which act over extremely large distances, and they nearly cancel for macroscopic objects. (Remember that it is the *net* external force that is important.) If they did not cancel, electromagnetic forces would completely overwhelm the gravitational force. The electromagnetic force is a combination of electrical forces (such as those that cause static electricity) and magnetic forces (such as those that affect a compass needle). These two forces were thought to be quite distinct until early in the 19th century, when scientists began to discover that they are different manifestations of the same force. This discovery is a classical case of the *unification of forces*. Similarly, friction, tension, and all of the other classes of forces we experience directly (except gravity, of course) are due to electromagnetic interactions of atoms and molecules. It is still convenient to consider these forces separately in specific applications, however, because of the ways they manifest themselves.

Note:

Concept Connections: Unifying Forces

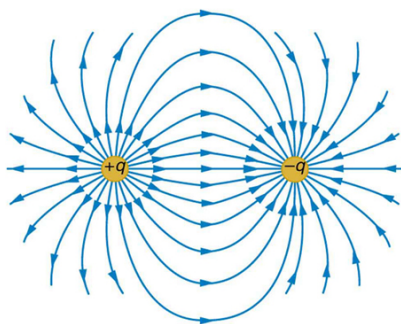
Attempts to unify the four basic forces are discussed in relation to elementary particles later in this text. By “unify” we mean finding connections between the forces that show that they are different manifestations of a single force. Even if such unification is achieved, the forces will retain their separate characteristics on the macroscopic scale and may be identical only under extreme conditions such as those existing in the early universe.

Physicists are now exploring whether the four basic forces are in some way related. Attempts to unify all forces into one come under the rubric of Grand Unified Theories (GUTs), with which there has been some success in recent years. It is now known that under conditions of extremely high density and temperature, such as existed in the early universe, the electromagnetic and weak nuclear forces are indistinguishable. They can now be considered to be different manifestations of one force, called the *electroweak* force. So the list of four has been reduced in a sense to only three. Further progress in unifying all forces is proving difficult—especially the inclusion of the gravitational force, which has the special characteristics of affecting the space and time in which the other forces exist.

While the unification of forces will not affect how we discuss forces in this text, it is fascinating that such underlying simplicity exists in the face of the overt complexity of the universe. There is no reason that nature must be simple—it simply is.

Action at a Distance: Concept of a Field

All forces act at a distance. This is obvious for the gravitational force. Earth and the Moon, for example, interact without coming into contact. It is also true for all other forces. Friction, for example, is an electromagnetic force between atoms that may not actually touch. What is it that carries forces between objects? One way to answer this question is to imagine that a **force field** surrounds whatever object creates the force. A second object (often called a *test object*) placed in this field will experience a force that is a function of location and other variables. The field itself is the “thing” that carries the force from one object to another. The field is defined so as to be a characteristic of the object creating it; the field does not depend on the test object placed in it. Earth’s gravitational field, for example, is a function of the mass of Earth and the distance from its center, independent of the presence of other masses. The concept of a field is useful because equations can be written for force fields surrounding objects (for gravity, this yields $w = mg$ at Earth’s surface), and motions can be calculated from these equations. (See [\[link\]](#).)



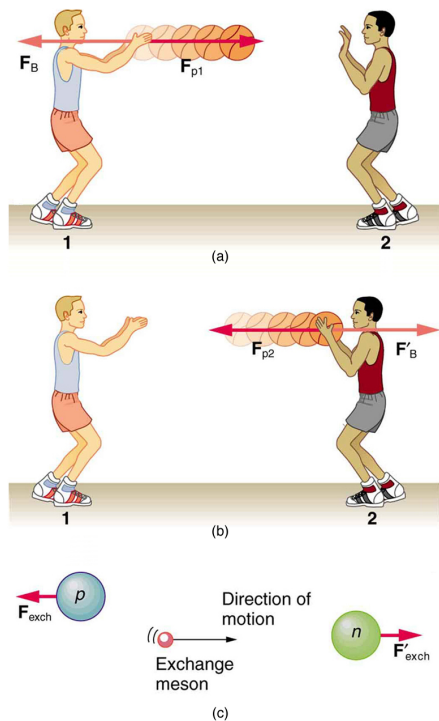
The electric force field between a positively charged particle and a negatively charged particle. When a positive test charge is placed in the field, the charge will experience a force in the direction of the force field lines.

Note:

Concept Connections: Force Fields

The concept of a *force field* is also used in connection with electric charge and is presented in [Electric Charge and Electric Field](#). It is also a useful idea for all the basic forces, as will be seen in [Particle Physics](#). Fields help us to visualize forces and how they are transmitted, as well as to describe them with precision and to link forces with subatomic carrier particles.

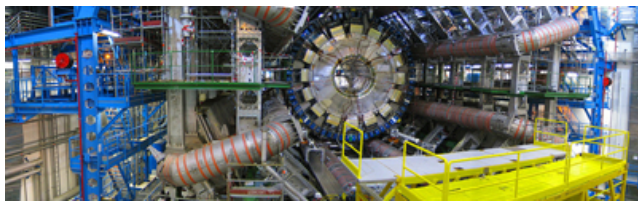
The field concept has been applied very successfully; we can calculate motions and describe nature to high precision using field equations. As useful as the field concept is, however, it leaves unanswered the question of what carries the force. It has been proposed in recent decades, starting in 1935 with Hideki Yukawa's (1907–1981) work on the strong nuclear force, that all forces are transmitted by the exchange of elementary particles. We can visualize particle exchange as analogous to macroscopic phenomena such as two people passing a basketball back and forth, thereby exerting a repulsive force without touching one another. (See [link](#).)



The exchange of masses resulting in repulsive forces.

(a) The person throwing the basketball exerts a force \mathbf{F}_{p1} on it toward the other person and feels a reaction force \mathbf{F}_B away from the second person. (b) The person catching the basketball exerts a force \mathbf{F}_{p2} on it to stop the ball and feels a reaction force \mathbf{F}'_B away from the first person. (c) The analogous exchange of a meson between a proton and a neutron carries the strong nuclear forces \mathbf{F}_{exch} and $\mathbf{F}'_{\text{exch}}$ between them. An attractive force can also be exerted by the exchange of a mass—if person 2 pulled the basketball away from the first person as he tried to retain it, then the force between them would be attractive.

This idea of particle exchange deepens rather than contradicts field concepts. It is more satisfying philosophically to think of something physical actually moving between objects acting at a distance. [\[link\]](#) lists the exchange or **carrier particles**, both observed and proposed, that carry the four forces. But the real fruit of the particle-exchange proposal is that searches for Yukawa's proposed particle found it *and* a number of others that were completely unexpected, stimulating yet more research. All of this research eventually led to the proposal of quarks as the underlying substructure of matter, which is a basic tenet of GUTs. If successful, these theories will explain not only forces, but also the structure of matter itself. Yet physics is an experimental science, so the test of these theories must lie in the domain of the real world. As of this writing, scientists at the CERN laboratory in Switzerland are starting to test these theories using the world's largest particle accelerator: the Large Hadron Collider. This accelerator (27 km in circumference) allows two high-energy proton beams, traveling in opposite directions, to collide. An energy of 14 trillion electron volts will be available. It is anticipated that some new particles, possibly force carrier particles, will be found. (See [\[link\]](#).) One of the force carriers of high interest that researchers hope to detect is the Higgs boson. The observation of its properties might tell us why different particles have different masses.



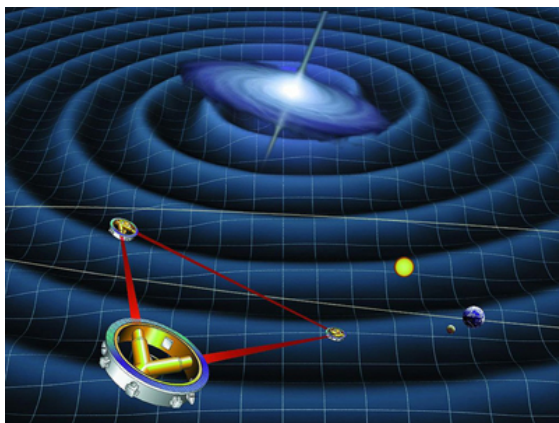
The world's largest particle accelerator spans the border between Switzerland and France. Two beams, traveling in opposite directions close to the speed of light, collide in a tube similar to the central tube shown here. External magnets determine the beam's path. Special detectors will analyze particles created in these collisions. Questions as broad as what is the origin of mass and what was matter like the first few seconds of our universe will be explored. This accelerator began preliminary operation in 2008. (credit: Frank Hommes)

Tiny particles also have wave-like behavior, something we will explore more in a later chapter. To better understand force-carrier particles from another perspective, let us consider gravity. The search for gravitational waves has been going on for a number of years. Almost 100 years

ago, Einstein predicted the existence of these waves as part of his general theory of relativity. Gravitational waves are created during the collision of massive stars, in black holes, or in supernova explosions—like shock waves. These gravitational waves will travel through space from such sites much like a pebble dropped into a pond sends out ripples—except these waves move at the speed of light. A detector apparatus has been built in the U.S., consisting of two large installations nearly 3000 km apart—one in Washington state and one in Louisiana! The facility is called the Laser Interferometer Gravitational-Wave Observatory (LIGO). Each installation is designed to use optical lasers to examine any slight shift in the relative positions of two masses due to the effect of gravity waves. The two sites allow simultaneous measurements of these small effects to be separated from other natural phenomena, such as earthquakes. Initial operation of the detectors began in 2002, and work is proceeding on increasing their sensitivity. Similar installations have been built in Italy (VIRGO), Germany (GEO600), and Japan (TAMA300) to provide a worldwide network of gravitational wave detectors.

International collaboration in this area is moving into space with the joint EU/US project LISA (Laser Interferometer Space Antenna). Earthquakes and other Earthly noises will be no problem for these monitoring spacecraft. LISA will complement LIGO by looking at much more massive black holes through the observation of gravitational-wave sources emitting much larger wavelengths. Three satellites will be placed in space above Earth in an equilateral triangle (with 5,000,000-km sides) ([\[link\]](#)). The system will measure the relative positions of each satellite to detect passing gravitational waves. Accuracy to within 10% of the size of an atom will be needed to detect any waves. The launch of this project might be as early as 2018.

“I’m sure LIGO will tell us something about the universe that we didn’t know before. The history of science tells us that any time you go where you haven’t been before, you usually find something that really shakes the scientific paradigms of the day. Whether gravitational wave astrophysics will do that, only time will tell.” —David Reitze, LIGO Input Optics Manager, University of Florida



Space-based future experiments for the measurement of gravitational waves. Shown here is a drawing of

LISA's orbit. Each satellite of LISA will consist of a laser source and a mass. The lasers will transmit a signal to measure the distance between each satellite's test mass. The relative motion of these masses will provide information about passing gravitational waves. (credit: NASA)

The ideas presented in this section are but a glimpse into topics of modern physics that will be covered in much greater depth in later chapters.

Summary

- The various types of forces that are categorized for use in many applications are all manifestations of the *four basic forces* in nature.
- The properties of these forces are summarized in [\[link\]](#).
- Everything we experience directly without sensitive instruments is due to either electromagnetic forces or gravitational forces. The nuclear forces are responsible for the submicroscopic structure of matter, but they are not directly sensed because of their short ranges. Attempts are being made to show all four forces are different manifestations of a single unified force.
- A force field surrounds an object creating a force and is the carrier of that force.

Conceptual Questions

Exercise:

Problem:

Explain, in terms of the properties of the four basic forces, why people notice the gravitational force acting on their bodies if it is such a comparatively weak force.

Exercise:

Problem:

What is the dominant force between astronomical objects? Why are the other three basic forces less significant over these very large distances?

Exercise:

Problem:

Give a detailed example of how the exchange of a particle can result in an *attractive* force. (For example, consider one child pulling a toy out of the hands of another.)

Problem Exercises

Exercise:

Problem:

(a) What is the strength of the weak nuclear force relative to the strong nuclear force? (b) What is the strength of the weak nuclear force relative to the electromagnetic force? Since the weak nuclear force acts at only very short distances, such as inside nuclei, where the strong and electromagnetic forces also act, it might seem surprising that we have any knowledge of it at all. We have such knowledge because the weak nuclear force is responsible for beta decay, a type of nuclear decay not explained by other forces.

Solution:

(a) 1×10^{-13}

(b) 1×10^{-11}

Exercise:

Problem:

(a) What is the ratio of the strength of the gravitational force to that of the strong nuclear force? (b) What is the ratio of the strength of the gravitational force to that of the weak nuclear force? (c) What is the ratio of the strength of the gravitational force to that of the electromagnetic force? What do your answers imply about the influence of the gravitational force on atomic nuclei?

Exercise:

Problem:

What is the ratio of the strength of the strong nuclear force to that of the electromagnetic force? Based on this ratio, you might expect that the strong force dominates the nucleus, which is true for small nuclei. Large nuclei, however, have sizes greater than the range of the strong nuclear force. At these sizes, the electromagnetic force begins to affect nuclear stability. These facts will be used to explain nuclear fusion and fission later in this text.

Solution:

10^2

Glossary

carrier particle

a fundamental particle of nature that is surrounded by a characteristic force field; photons are carrier particles of the electromagnetic force

force field

a region in which a test particle will experience a force

Introduction to Uniform Circular Motion and Gravitation

class="introduction"

This
Australian
Grand Prix
Formula 1
race car
moves in a
circular
path as it
makes the
turn. Its
wheels also
spin rapidly
—the latter
completing
many
revolutions,
the former
only part of
one (a
circular
arc). The
same
physical
principles
are
involved in
each.
(credit:
Richard
Munckton)



Many motions, such as the arc of a bird's flight or Earth's path around the Sun, are curved. Recall that Newton's first law tells us that motion is along a straight line at constant speed unless there is a net external force. We will therefore study not only motion along curves, but also the forces that cause it, including gravitational forces. In some ways, this chapter is a continuation of [Dynamics: Newton's Laws of Motion](#) as we study more applications of Newton's laws of motion.

This chapter deals with the simplest form of curved motion, **uniform circular motion**, motion in a circular path at constant speed. Studying this topic illustrates most concepts associated with rotational motion and leads to the study of many new topics we group under the name *rotation*. Pure *rotational motion* occurs when points in an object move in circular paths centered on one point. Pure *translational motion* is motion with no rotation. Some motion combines both types, such as a rotating hockey puck moving along ice.

Glossary

uniform circular motion

the motion of an object in a circular path at constant speed

Rotation Angle and Angular Velocity

- Define arc length, rotation angle, radius of curvature and angular velocity.
- Calculate the angular velocity of a car wheel spin.

In [Kinematics](#), we studied motion along a straight line and introduced such concepts as displacement, velocity, and acceleration. [Two-Dimensional Kinematics](#) dealt with motion in two dimensions. Projectile motion is a special case of two-dimensional kinematics in which the object is projected into the air, while being subject to the gravitational force, and lands a distance away. In this chapter, we consider situations where the object does not land but moves in a curve. We begin the study of uniform circular motion by defining two angular quantities needed to describe rotational motion.

Rotation Angle

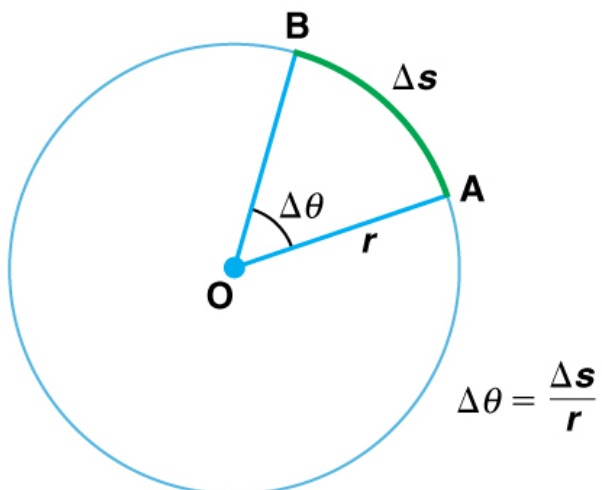
When objects rotate about some axis—for example, when the CD (compact disc) in [\[link\]](#) rotates about its center—each point in the object follows a circular arc. Consider a line from the center of the CD to its edge. Each **pit** used to record sound along this line moves through the same angle in the same amount of time. The rotation angle is the amount of rotation and is analogous to linear distance. We define the **rotation angle** $\Delta\theta$ to be the ratio of the arc length to the radius of curvature:

Equation:

$$\Delta\theta = \frac{\Delta s}{r}.$$



All points on a CD travel in circular arcs. The pits along a line from the center to the edge all move through the same angle $\Delta\theta$ in a time Δt .



The radius of a circle is rotated through an angle $\Delta\theta$. The arc

length Δs is described on the circumference.

The **arc length** Δs is the distance traveled along a circular path as shown in [\[link\]](#) Note that r is the **radius of curvature** of the circular path.

We know that for one complete revolution, the arc length is the circumference of a circle of radius r . The circumference of a circle is $2\pi r$. Thus for one complete revolution the rotation angle is

Equation:

$$\Delta\theta = \frac{2\pi r}{r} = 2\pi.$$

This result is the basis for defining the units used to measure rotation angles, $\Delta\theta$ to be **radians** (rad), defined so that

Equation:

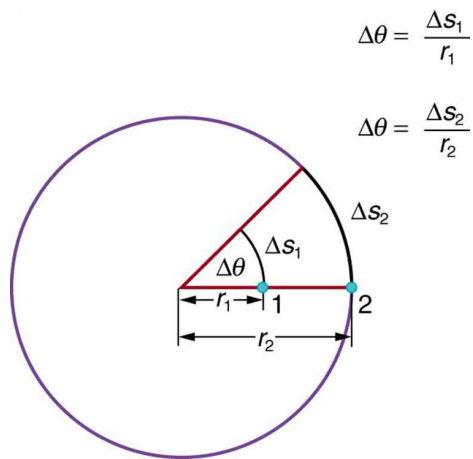
$$2\pi \text{ rad} = 1 \text{ revolution.}$$

A comparison of some useful angles expressed in both degrees and radians is shown in [\[link\]](#).

Degree Measures	Radian Measure
30°	$\frac{\pi}{6}$

Degree Measures	Radian Measure
60°	$\frac{\pi}{3}$
90°	$\frac{\pi}{2}$
120°	$\frac{2\pi}{3}$
135°	$\frac{3\pi}{4}$
180°	π

Comparison of Angular Units



Points 1 and 2 rotate through the same angle ($\Delta\theta$), but point 2 moves through a greater arc length (Δs) because it is at a greater distance from the center of rotation (r).

If $\Delta\theta = 2\pi$ rad, then the CD has made one complete revolution, and every point on the CD is back at its original position. Because there are 360° in a circle or one revolution, the relationship between radians and degrees is thus

Equation:

$$2\pi \text{ rad} = 360^\circ$$

so that

Equation:

$$1 \text{ rad} = \frac{360^\circ}{2\pi} \approx 57.3^\circ.$$

Angular Velocity

How fast is an object rotating? We define **angular velocity** ω as the rate of change of an angle. In symbols, this is

Equation:

$$\omega = \frac{\Delta\theta}{\Delta t},$$

where an angular rotation $\Delta\theta$ takes place in a time Δt . The greater the rotation angle in a given amount of time, the greater the angular velocity. The units for angular velocity are radians per second (rad/s).

Angular velocity ω is analogous to linear velocity v . To get the precise relationship between angular and linear velocity, we again consider a pit on the rotating CD. This pit moves an arc length Δs in a time Δt , and so it has a linear velocity

Equation:

$$v = \frac{\Delta s}{\Delta t}.$$

From $\Delta\theta = \frac{\Delta s}{r}$ we see that $\Delta s = r\Delta\theta$. Substituting this into the expression for v gives

Equation:

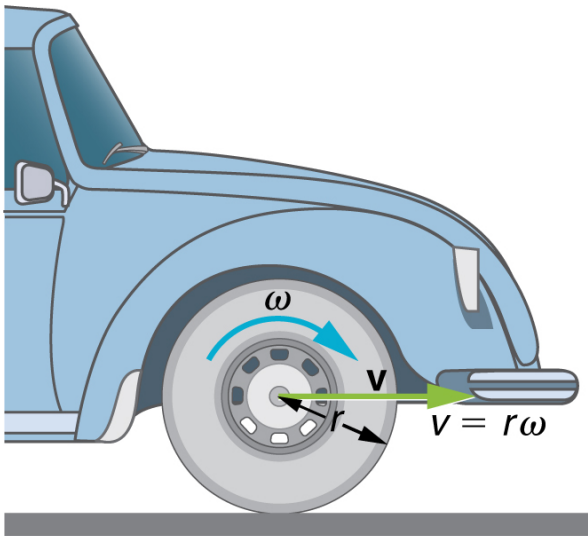
$$v = \frac{r\Delta\theta}{\Delta t} = r\omega.$$

We write this relationship in two different ways and gain two different insights:

Equation:

$$v = r\omega \text{ or } \omega = \frac{v}{r}.$$

The first relationship in $v = r\omega$ or $\omega = \frac{v}{r}$ states that the linear velocity v is proportional to the distance from the center of rotation, thus, it is largest for a point on the rim (largest r), as you might expect. We can also call this linear speed v of a point on the rim the *tangential speed*. The second relationship in $v = r\omega$ or $\omega = \frac{v}{r}$ can be illustrated by considering the tire of a moving car. Note that the speed of a point on the rim of the tire is the same as the speed v of the car. See [\[link\]](#). So the faster the car moves, the faster the tire spins—large v means a large ω , because $v = r\omega$. Similarly, a larger-radius tire rotating at the same angular velocity (ω) will produce a greater linear speed (v) for the car.



A car moving at a velocity v to the right has a tire rotating with an angular velocity ω . The speed of the tread of the tire relative to the axle is v , the same as if the car were jacked up. Thus the car moves forward at linear velocity $v = r\omega$, where r is the tire radius. A larger angular velocity for the tire means a greater velocity for the car.

Example:**How Fast Does a Car Tire Spin?**

Calculate the angular velocity of a 0.300 m radius car tire when the car travels at 15.0 m/s (about 54 km/h). See [\[link\]](#).

Strategy

Because the linear speed of the tire rim is the same as the speed of the car, we have $v = 15.0 \text{ m/s}$. The radius of the tire is given to be $r = 0.300 \text{ m}$. Knowing v and r , we can use the second relationship in $v = r\omega$, $\omega = \frac{v}{r}$ to calculate the angular velocity.

Solution

To calculate the angular velocity, we will use the following relationship:

Equation:

$$\omega = \frac{v}{r}.$$

Substituting the knowns,

Equation:

$$\omega = \frac{15.0 \text{ m/s}}{0.300 \text{ m}} = 50.0 \text{ rad/s}.$$

Discussion

When we cancel units in the above calculation, we get 50.0/s. But the angular velocity must have units of rad/s. Because radians are actually unitless (radians are defined as a ratio of distance), we can simply insert them into the answer for the angular velocity. Also note that if an earth mover with much larger tires, say 1.20 m in radius, were moving at the same speed of 15.0 m/s, its tires would rotate more slowly. They would have an angular velocity

Equation:

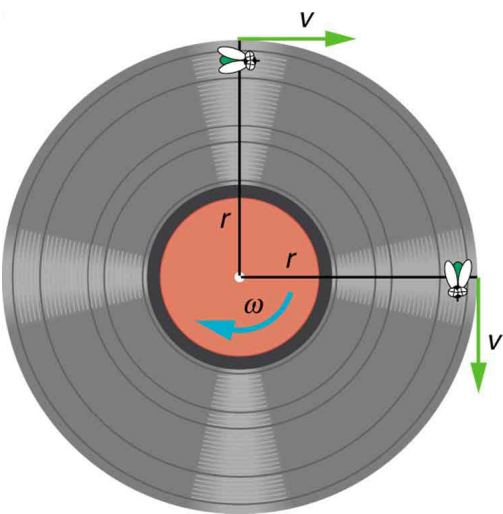
$$\omega = (15.0 \text{ m/s})/(1.20 \text{ m}) = 12.5 \text{ rad/s}.$$

Both ω and v have directions (hence they are angular and linear *velocities*, respectively). Angular velocity has only two directions with respect to the axis of rotation—it is either clockwise or counterclockwise. Linear velocity is tangent to the path, as illustrated in [\[link\]](#).

Note:

Take-Home Experiment

Tie an object to the end of a string and swing it around in a horizontal circle above your head (swing at your wrist). Maintain uniform speed as the object swings and measure the angular velocity of the motion. What is the approximate speed of the object? Identify a point close to your hand and take appropriate measurements to calculate the linear speed at this point. Identify other circular motions and measure their angular velocities.



As an object moves in a circle, here a fly on the edge of an old-fashioned vinyl record, its instantaneous velocity is

always tangent to the circle. The direction of the angular velocity is clockwise in this case.

Note:

PhET Explorations: Ladybug Revolution

[Ladybug
Revolutio
n](#)

Join the ladybug in an exploration of rotational motion. Rotate the merry-go-round to change its angle, or choose a constant angular velocity or angular acceleration. Explore how circular motion relates to the bug's x,y position, velocity, and acceleration using vectors or graphs.

Section Summary

- Uniform circular motion is motion in a circle at constant speed. The rotation angle $\Delta\theta$ is defined as the ratio of the arc length to the radius of curvature:

Equation:

$$\Delta\theta = \frac{\Delta s}{r},$$

where arc length Δs is distance traveled along a circular path and r is the radius of curvature of the circular path. The quantity $\Delta\theta$ is

measured in units of radians (rad), for which

Equation:

$$2\pi \text{ rad} = 360^\circ = 1 \text{ revolution.}$$

- The conversion between radians and degrees is $1 \text{ rad} = 57.3^\circ$.
- Angular velocity ω is the rate of change of an angle,

Equation:

$$\omega = \frac{\Delta\theta}{\Delta t},$$

where a rotation $\Delta\theta$ takes place in a time Δt . The units of angular velocity are radians per second (rad/s). Linear velocity v and angular velocity ω are related by

Equation:

$$v = r\omega \text{ or } \omega = \frac{v}{r}.$$

Conceptual Questions

Exercise:

Problem:

There is an analogy between rotational and linear physical quantities. What rotational quantities are analogous to distance and velocity?

Problem Exercises

Exercise:

Problem:

Semi-trailer trucks have an odometer on one hub of a trailer wheel. The hub is weighted so that it does not rotate, but it contains gears to count the number of wheel revolutions—it then calculates the distance traveled. If the wheel has a 1.15 m diameter and goes through 200,000 rotations, how many kilometers should the odometer read?

Solution:

723 km

Exercise:**Problem:**

Microwave ovens rotate at a rate of about 6 rev/min. What is this in revolutions per second? What is the angular velocity in radians per second?

Exercise:**Problem:**

An automobile with 0.260 m radius tires travels 80,000 km before wearing them out. How many revolutions do the tires make, neglecting any backing up and any change in radius due to wear?

Solution:

5×10^7 rotations

Exercise:**Problem:**

(a) What is the period of rotation of Earth in seconds? (b) What is the angular velocity of Earth? (c) Given that Earth has a radius of 6.4×10^6 m at its equator, what is the linear velocity at Earth's surface?

Exercise:**Problem:**

A baseball pitcher brings his arm forward during a pitch, rotating the forearm about the elbow. If the velocity of the ball in the pitcher's hand is 35.0 m/s and the ball is 0.300 m from the elbow joint, what is the angular velocity of the forearm?

Solution:

117 rad/s

Exercise:**Problem:**

In lacrosse, a ball is thrown from a net on the end of a stick by rotating the stick and forearm about the elbow. If the angular velocity of the ball about the elbow joint is 30.0 rad/s and the ball is 1.30 m from the elbow joint, what is the velocity of the ball?

Exercise:**Problem:**

A truck with 0.420-m -radius tires travels at 32.0 m/s . What is the angular velocity of the rotating tires in radians per second? What is this in rev/min ?

Solution:

76.2 rad/s

728 rpm

Exercise:**Problem:**

Integrated Concepts When kicking a football, the kicker rotates his leg about the hip joint.

- (a) If the velocity of the tip of the kicker's shoe is 35.0 m/s and the hip joint is 1.05 m from the tip of the shoe, what is the shoe tip's angular velocity?
- (b) The shoe is in contact with the initially stationary 0.500 kg football for 20.0 ms. What average force is exerted on the football to give it a velocity of 20.0 m/s?
- (c) Find the maximum range of the football, neglecting air resistance.
-

Solution:

- (a) 33.3 rad/s
- (b) 500 N
- (c) 40.8 m

Exercise:

Problem:Construct Your Own Problem

Consider an amusement park ride in which participants are rotated about a vertical axis in a cylinder with vertical walls. Once the angular velocity reaches its full value, the floor drops away and friction between the walls and the riders prevents them from sliding down. Construct a problem in which you calculate the necessary angular velocity that assures the riders will not slide down the wall. Include a free body diagram of a single rider. Among the variables to consider are the radius of the cylinder and the coefficients of friction between the riders' clothing and the wall.

Glossary

arc length

Δs , the distance traveled by an object along a circular path

pit

a tiny indentation on the spiral track moulded into the top of the polycarbonate layer of CD

rotation angle

the ratio of the arc length to the radius of curvature on a circular path:

$$\Delta\theta = \frac{\Delta s}{r}$$

radius of curvature

radius of a circular path

radians

a unit of angle measurement

angular velocity

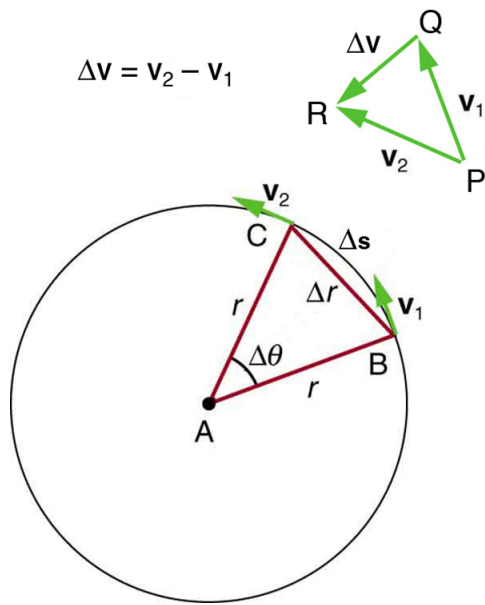
ω , the rate of change of the angle with which an object moves on a circular path

Centripetal Acceleration

- Establish the expression for centripetal acceleration.
- Explain the centrifuge.

We know from kinematics that acceleration is a change in velocity, either in its magnitude or in its direction, or both. In uniform circular motion, the direction of the velocity changes constantly, so there is always an associated acceleration, even though the magnitude of the velocity might be constant. You experience this acceleration yourself when you turn a corner in your car. (If you hold the wheel steady during a turn and move at constant speed, you are in uniform circular motion.) What you notice is a sideways acceleration because you and the car are changing direction. The sharper the curve and the greater your speed, the more noticeable this acceleration will become. In this section we examine the direction and magnitude of that acceleration.

[\[link\]](#) shows an object moving in a circular path at constant speed. The direction of the instantaneous velocity is shown at two points along the path. Acceleration is in the direction of the change in velocity, which points directly toward the center of rotation (the center of the circular path). This pointing is shown with the vector diagram in the figure. We call the acceleration of an object moving in uniform circular motion (resulting from a net external force) the **centripetal acceleration**(a_c); centripetal means “toward the center” or “center seeking.”



The directions of the velocity of an object at two different points are shown, and the change in velocity $\Delta\mathbf{v}$ is seen to point directly toward the center of curvature. (See small inset.) Because $\mathbf{a}_c = \Delta\mathbf{v}/\Delta t$, the acceleration is also toward the center; \mathbf{a}_c is called centripetal acceleration. (Because $\Delta\theta$ is very small, the arc length Δs is equal to the chord length Δr for small time differences.)

The direction of centripetal acceleration is toward the center of curvature, but what is its magnitude? Note that the triangle formed by the velocity vectors and the one formed by the radii r and Δs are similar. Both the

triangles ABC and PQR are isosceles triangles (two equal sides). The two equal sides of the velocity vector triangle are the speeds $v_1 = v_2 = v$. Using the properties of two similar triangles, we obtain

Equation:

$$\frac{\Delta v}{v} = \frac{\Delta s}{r}.$$

Acceleration is $\Delta v / \Delta t$, and so we first solve this expression for Δv :

Equation:

$$\Delta v = \frac{v}{r} \Delta s.$$

Then we divide this by Δt , yielding

Equation:

$$\frac{\Delta v}{\Delta t} = \frac{v}{r} \times \frac{\Delta s}{\Delta t}.$$

Finally, noting that $\Delta v / \Delta t = a_c$ and that $\Delta s / \Delta t = v$, the linear or tangential speed, we see that the magnitude of the centripetal acceleration is

Equation:

$$a_c = \frac{v^2}{r},$$

which is the acceleration of an object in a circle of radius r at a speed v . So, centripetal acceleration is greater at high speeds and in sharp curves (smaller radius), as you have noticed when driving a car. But it is a bit surprising that a_c is proportional to speed squared, implying, for example, that it is four times as hard to take a curve at 100 km/h than at 50 km/h. A sharp corner has a small radius, so that a_c is greater for tighter turns, as you have probably noticed.

It is also useful to express a_c in terms of angular velocity. Substituting $v = r\omega$ into the above expression, we find $a_c = (r\omega)^2/r = r\omega^2$. We can express the magnitude of centripetal acceleration using either of two equations:

Equation:

$$a_c = \frac{v^2}{r}; \quad a_c = r\omega^2.$$

Recall that the direction of a_c is toward the center. You may use whichever expression is more convenient, as illustrated in examples below.

A **centrifuge** (see [\[link\]](#)b) is a rotating device used to separate specimens of different densities. High centripetal acceleration significantly decreases the time it takes for separation to occur, and makes separation possible with small samples. Centrifuges are used in a variety of applications in science and medicine, including the separation of single cell suspensions such as bacteria, viruses, and blood cells from a liquid medium and the separation of macromolecules, such as DNA and protein, from a solution. Centrifuges are often rated in terms of their centripetal acceleration relative to acceleration due to gravity (g); maximum centripetal acceleration of several hundred thousand g is possible in a vacuum. Human centrifuges, extremely large centrifuges, have been used to test the tolerance of astronauts to the effects of accelerations larger than that of Earth's gravity.

Example:

How Does the Centripetal Acceleration of a Car Around a Curve Compare with That Due to Gravity?

What is the magnitude of the centripetal acceleration of a car following a curve of radius 500 m at a speed of 25.0 m/s (about 90 km/h)? Compare the acceleration with that due to gravity for this fairly gentle curve taken at highway speed. See [\[link\]](#)(a).

Strategy

Because v and r are given, the first expression in $a_c = \frac{v^2}{r}$; $a_c = r\omega^2$ is the most convenient to use.

Solution

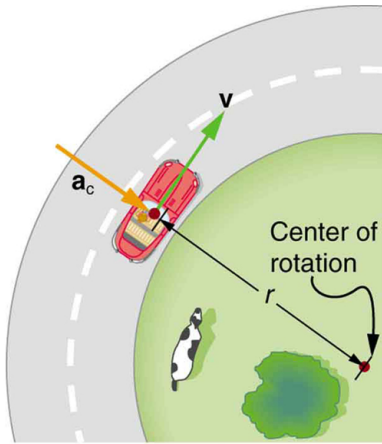
Entering the given values of $v = 25.0 \text{ m/s}$ and $r = 500 \text{ m}$ into the first expression for a_c gives

Equation:

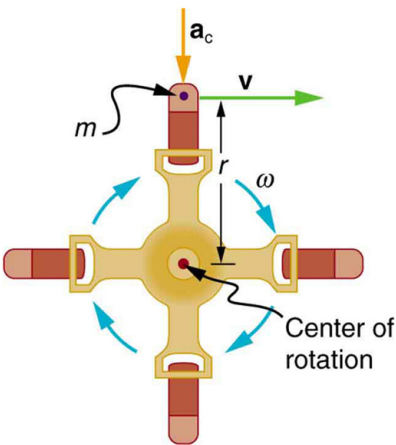
$$a_c = \frac{v^2}{r} = \frac{(25.0 \text{ m/s})^2}{500 \text{ m}} = 1.25 \text{ m/s}^2.$$

Discussion

To compare this with the acceleration due to gravity ($g = 9.80 \text{ m/s}^2$), we take the ratio of $a_c/g = (1.25 \text{ m/s}^2) / (9.80 \text{ m/s}^2) = 0.128$. Thus, $a_c = 0.128 g$ and is noticeable especially if you were not wearing a seat belt.



(a) Car around corner



(b) Centrifuge

(a) The car following a circular path at constant speed is accelerated perpendicular to its velocity, as shown. The magnitude of this centripetal acceleration is found in [\[link\]](#). (b) A particle of mass in a centrifuge is rotating at constant angular velocity . It

must be accelerated perpendicular to its velocity or it would continue in a straight line. The magnitude of the necessary acceleration is found in [\[link\]](#).

Example:

How Big Is the Centripetal Acceleration in an Ultracentrifuge?

Calculate the centripetal acceleration of a point 7.50 cm from the axis of an **ultracentrifuge** spinning at 7.5×10^4 rev/min. Determine the ratio of this acceleration to that due to gravity. See [\[link\]](#)(b).

Strategy

The term rev/min stands for revolutions per minute. By converting this to radians per second, we obtain the angular velocity ω . Because r is given, we can use the second expression in the equation $a_c = \frac{v^2}{r}$; $a_c = r\omega^2$ to calculate the centripetal acceleration.

Solution

To convert 7.50×10^4 rev/min to radians per second, we use the facts that one revolution is 2π rad and one minute is 60.0 s. Thus,

Equation:

$$\omega = 7.50 \times 10^4 \frac{\text{rev}}{\text{min}} \times \frac{2\pi \text{ rad}}{1 \text{ rev}} \times \frac{1 \text{ min}}{60.0 \text{ s}} = 7854 \text{ rad/s}.$$

Now the centripetal acceleration is given by the second expression in

$$a_c = \frac{v^2}{r}; a_c = r\omega^2 \text{ as}$$

Equation:

$$a_c = r\omega^2.$$

Converting 7.50 cm to meters and substituting known values gives

Equation:

$$a_c = (0.0750 \text{ m})(7854 \text{ rad/s})^2 = 4.63 \times 10^6 \text{ m/s}^2.$$

Note that the unitless radians are discarded in order to get the correct units for centripetal acceleration. Taking the ratio of a_c to g yields

Equation:

$$\frac{a_c}{g} = \frac{4.63 \times 10^6}{9.80} = 4.72 \times 10^5.$$

Discussion

This last result means that the centripetal acceleration is 472,000 times as strong as g . It is no wonder that such high ω centrifuges are called ultracentrifuges. The extremely large accelerations involved greatly decrease the time needed to cause the sedimentation of blood cells or other materials.

Of course, a net external force is needed to cause any acceleration, just as Newton proposed in his second law of motion. So a net external force is needed to cause a centripetal acceleration. In [Centripetal Force](#), we will consider the forces involved in circular motion.

Note:

PhET Explorations: Ladybug Motion 2D

Learn about position, velocity and acceleration vectors. Move the ladybug by setting the position, velocity or acceleration, and see how the vectors change. Choose linear, circular or elliptical motion, and record and playback the motion to analyze the behavior.

<https://archive.cnx.org/specials/317a2b1e-2fbd-11e5-99b5-e38ffb545fe6/ladybug-motion/#sim-ladybug-motion>

Section Summary

- Centripetal acceleration a_c is the acceleration experienced while in uniform circular motion. It always points toward the center of rotation. It is perpendicular to the linear velocity v and has the magnitude
Equation:

$$a_c = \frac{v^2}{r}; a_c = r\omega^2.$$

- The unit of centripetal acceleration is m/s^2 .

Conceptual Questions

Exercise:

Problem:

Can centripetal acceleration change the speed of circular motion? Explain.

Problem Exercises

Exercise:

Problem:

A fairground ride spins its occupants inside a flying saucer-shaped container. If the horizontal circular path the riders follow has an 8.00 m radius, at how many revolutions per minute will the riders be subjected to a centripetal acceleration whose magnitude is 1.50 times that due to gravity?

Solution:

12.9 rev/min

Exercise:

Problem:

A runner taking part in the 200 m dash must run around the end of a track that has a circular arc with a radius of curvature of 30 m. If he completes the 200 m dash in 23.2 s and runs at constant speed throughout the race, what is the magnitude of his centripetal acceleration as he runs the curved portion of the track?

Exercise:**Problem:**

Taking the age of Earth to be about 4×10^9 years and assuming its orbital radius of 1.5×10^{11} m has not changed and is circular, calculate the approximate total distance Earth has traveled since its birth (in a frame of reference stationary with respect to the Sun).

Solution:

$$4 \times 10^{21} \text{ m}$$

Exercise:**Problem:**

The propeller of a World War II fighter plane is 2.30 m in diameter.

(a) What is its angular velocity in radians per second if it spins at 1200 rev/min?

(b) What is the linear speed of its tip at this angular velocity if the plane is stationary on the tarmac?

(c) What is the centripetal acceleration of the propeller tip under these conditions? Calculate it in meters per second squared and convert to multiples of g .

Exercise:

Problem:

An ordinary workshop grindstone has a radius of 7.50 cm and rotates at 6500 rev/min.

- (a) Calculate the magnitude of the centripetal acceleration at its edge in meters per second squared and convert it to multiples of g .
 - (b) What is the linear speed of a point on its edge?
-

Solution:

- a) $3.47 \times 10^4 \text{ m/s}^2$, $3.55 \times 10^3 g$
- b) 51.1 m/s

Exercise:**Problem:**

Helicopter blades withstand tremendous stresses. In addition to supporting the weight of a helicopter, they are spun at rapid rates and experience large centripetal accelerations, especially at the tip.

- (a) Calculate the magnitude of the centripetal acceleration at the tip of a 4.00 m long helicopter blade that rotates at 300 rev/min.
- (b) Compare the linear speed of the tip with the speed of sound (taken to be 340 m/s).

Exercise:

Problem: Olympic ice skaters are able to spin at about 5 rev/s.

- (a) What is their angular velocity in radians per second?
- (b) What is the centripetal acceleration of the skater's nose if it is 0.120 m from the axis of rotation?

(c) An exceptional skater named Dick Button was able to spin much faster in the 1950s than anyone since—at about 9 rev/s. What was the centripetal acceleration of the tip of his nose, assuming it is at 0.120 m radius?

(d) Comment on the magnitudes of the accelerations found. It is reputed that Button ruptured small blood vessels during his spins.

Solution:

a) 31.4 rad/s

b) 118 m/s

c) 384 m/s

d) The centripetal acceleration felt by Olympic skaters is 12 times larger than the acceleration due to gravity. That's quite a lot of acceleration in itself. The centripetal acceleration felt by Button's nose was 39.2 times larger than the acceleration due to gravity. It is no wonder that he ruptured small blood vessels in his spins.

Exercise:

Problem:

What percentage of the acceleration at Earth's surface is the acceleration due to gravity at the position of a satellite located 300 km above Earth?

Exercise:

Problem:

Verify that the linear speed of an ultracentrifuge is about 0.50 km/s, and Earth in its orbit is about 30 km/s by calculating:

(a) The linear speed of a point on an ultracentrifuge 0.100 m from its center, rotating at 50,000 rev/min.

(b) The linear speed of Earth in its orbit about the Sun (use data from the text on the radius of Earth's orbit and approximate it as being circular).

Solution:

a) 0.524 km/s

b) 29.7 km/s

Exercise:

Problem:

A rotating space station is said to create “artificial gravity”—a loosely-defined term used for an acceleration that would be crudely similar to gravity. The outer wall of the rotating space station would become a floor for the astronauts, and centripetal acceleration supplied by the floor would allow astronauts to exercise and maintain muscle and bone strength more naturally than in non-rotating space environments. If the space station is 200 m in diameter, what angular velocity would produce an “artificial gravity” of 9.80 m/s^2 at the rim?

Exercise:

Problem:

At takeoff, a commercial jet has a 60.0 m/s speed. Its tires have a diameter of 0.850 m.

(a) At how many rev/min are the tires rotating?

(b) What is the centripetal acceleration at the edge of the tire?

(c) With what force must a determined $1.00 \times 10^{-15} \text{ kg}$ bacterium cling to the rim?

(d) Take the ratio of this force to the bacterium's weight.

Solution:

(a) 1.35×10^3 rpm

(b) 8.47×10^3 m/s²

(c) 8.47×10^{-12} N

(d) 865

Exercise:

Problem: Integrated Concepts

Riders in an amusement park ride shaped like a Viking ship hung from a large pivot are rotated back and forth like a rigid pendulum. Sometime near the middle of the ride, the ship is momentarily motionless at the top of its circular arc. The ship then swings down under the influence of gravity.

(a) Assuming negligible friction, find the speed of the riders at the bottom of its arc, given the system's center of mass travels in an arc having a radius of 14.0 m and the riders are near the center of mass.

(b) What is the centripetal acceleration at the bottom of the arc?

(c) Draw a free body diagram of the forces acting on a rider at the bottom of the arc.

(d) Find the force exerted by the ride on a 60.0 kg rider and compare it to her weight.

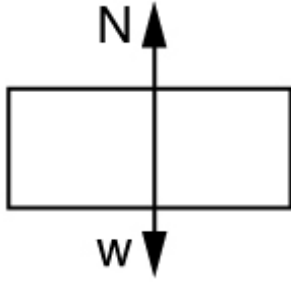
(e) Discuss whether the answer seems reasonable.

Solution:

(a) 16.6 m/s

(b) 19.6 m/s²

(c)



(d) $1.76 \times 10^3 \text{ N}$ or $3.00 w$, that is, the normal force (upward) is three times her weight.

(e) This answer seems reasonable, since she feels like she's being forced into the chair MUCH stronger than just by gravity.

Exercise:

Problem: Unreasonable Results

A mother pushes her child on a swing so that his speed is 9.00 m/s at the lowest point of his path. The swing is suspended 2.00 m above the child's center of mass.

(a) What is the magnitude of the centripetal acceleration of the child at the low point?

(b) What is the magnitude of the force the child exerts on the seat if his mass is 18.0 kg ?

(c) What is unreasonable about these results?

(d) Which premises are unreasonable or inconsistent?

Solution:

a) 40.5 m/s^2

b) 905 N

c) The force in part (b) is very large. The acceleration in part (a) is too much, about 4 g.

d) The speed of the swing is too large. At the given velocity at the bottom of the swing, there is enough kinetic energy to send the child all the way over the top, ignoring friction.

Glossary

centripetal acceleration

the acceleration of an object moving in a circle, directed toward the center

ultracentrifuge

a centrifuge optimized for spinning a rotor at very high speeds

Centripetal Force

- Calculate coefficient of friction on a car tire.
- Calculate ideal speed and angle of a car on a turn.

Any force or combination of forces can cause a centripetal or radial acceleration. Just a few examples are the tension in the rope on a tether ball, the force of Earth's gravity on the Moon, friction between roller skates and a rink floor, a banked roadway's force on a car, and forces on the tube of a spinning centrifuge.

Any net force causing uniform circular motion is called a **centripetal force**. The direction of a centripetal force is toward the center of curvature, the same as the direction of centripetal acceleration. According to Newton's second law of motion, net force is mass times acceleration: net $F = ma$. For uniform circular motion, the acceleration is the centripetal acceleration — $a = a_c$. Thus, the magnitude of centripetal force F_c is

Equation:

$$F_c = ma_c.$$

By using the expressions for centripetal acceleration a_c from $a_c = \frac{v^2}{r}$; $a_c = r\omega^2$, we get two expressions for the centripetal force F_c in terms of mass, velocity, angular velocity, and radius of curvature:

Equation:

$$F_c = m\frac{v^2}{r}; F_c = mr\omega^2.$$

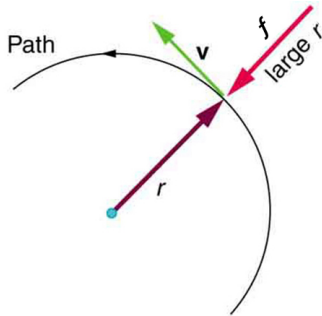
You may use whichever expression for centripetal force is more convenient. Centripetal force F_c is always perpendicular to the path and pointing to the center of curvature, because \mathbf{a}_c is perpendicular to the velocity and pointing to the center of curvature.

Note that if you solve the first expression for r , you get

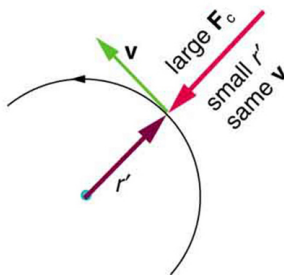
Equation:

$$r = \frac{mv^2}{F_c}.$$

This implies that for a given mass and velocity, a large centripetal force causes a small radius of curvature—that is, a tight curve.



$f = F_c$ is parallel to a_c since $F_c = ma_c$



The frictional force
supplies the
centripetal force
and is numerically
equal to it.

Centripetal force is
perpendicular to
velocity and causes
uniform circular
motion. The larger
the F_c , the smaller
the radius of
curvature r and the
sharper the curve.
The second curve

has the same v , but
a larger F_c
produces a smaller
 r !

Example:

What Coefficient of Friction Do Car Tires Need on a Flat Curve?

(a) Calculate the centripetal force exerted on a 900 kg car that negotiates a 500 m radius curve at 25.0 m/s.

(b) Assuming an unbanked curve, find the minimum static coefficient of friction, between the tires and the road, static friction being the reason that keeps the car from slipping (see [\[link\]](#)).

Strategy and Solution for (a)

We know that $F_c = \frac{mv^2}{r}$. Thus,

Equation:

$$F_c = \frac{mv^2}{r} = \frac{(900 \text{ kg})(25.0 \text{ m/s})^2}{(500 \text{ m})} = 1125 \text{ N}.$$

Strategy for (b)

[\[link\]](#) shows the forces acting on the car on an unbanked (level ground) curve. Friction is to the left, keeping the car from slipping, and because it is the only horizontal force acting on the car, the friction is the centripetal force in this case. We know that the maximum static friction (at which the tires roll but do not slip) is $\mu_s N$, where μ_s is the static coefficient of friction and N is the normal force. The normal force equals the car's weight on level ground, so that $N = mg$. Thus the centripetal force in this situation is

Equation:

$$F_c = f = \mu_s N = \mu_s mg.$$

Now we have a relationship between centripetal force and the coefficient of friction. Using the first expression for F_c from the equation

Equation:

$$\left. \begin{aligned} F_c &= m \frac{v^2}{r} \\ F_c &= mr\omega^2 \end{aligned} \right\},$$

Equation:

$$m \frac{v^2}{r} = \mu_s mg.$$

We solve this for μ_s , noting that mass cancels, and obtain

Equation:

$$\mu_s = \frac{v^2}{rg}.$$

Solution for (b)

Substituting the knowns,

Equation:

$$\mu_s = \frac{(25.0 \text{ m/s})^2}{(500 \text{ m})(9.80 \text{ m/s}^2)} = 0.13.$$

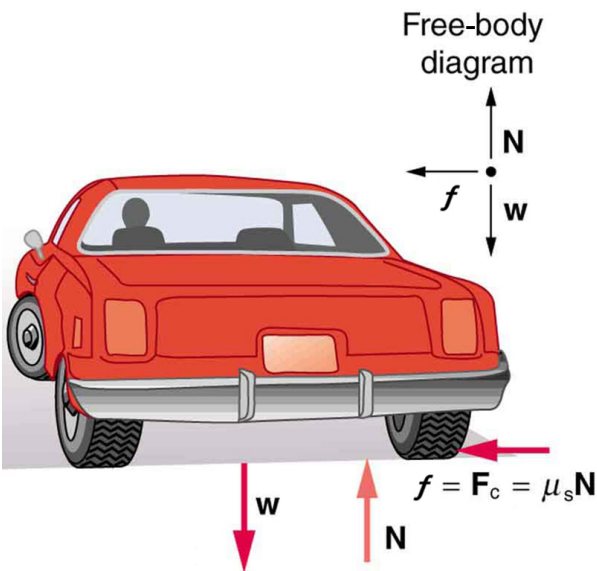
(Because coefficients of friction are approximate, the answer is given to only two digits.)

Discussion

We could also solve part (a) using the first expression in $\left. \begin{aligned} F_c &= m \frac{v^2}{r} \\ F_c &= mr\omega^2 \end{aligned} \right\},$

because m, v , and r are given. The coefficient of friction found in part (b) is much smaller than is typically found between tires and roads. The car will still negotiate the curve if the coefficient is greater than 0.13, because static friction is a responsive force, being able to assume a value less than but no more than $\mu_s N$. A higher coefficient would also allow the car to negotiate the curve at a higher speed, but if the coefficient of friction is

less, the safe speed would be less than 25 m/s. Note that mass cancels, implying that in this example, it does not matter how heavily loaded the car is to negotiate the turn. Mass cancels because friction is assumed proportional to the normal force, which in turn is proportional to mass. If the surface of the road were banked, the normal force would be less as will be discussed below.



This car on level ground is moving away and turning to the left. The centripetal force causing the car to turn in a circular path is due to friction between the tires and the road.

A minimum coefficient of friction is needed, or the car will move in a larger-radius curve and leave the roadway.

Let us now consider **banked curves**, where the slope of the road helps you negotiate the curve. See [\[link\]](#). The greater the angle θ , the faster you can

take the curve. Race tracks for bikes as well as cars, for example, often have steeply banked curves. In an “ideally banked curve,” the angle θ is such that you can negotiate the curve at a certain speed without the aid of friction between the tires and the road. We will derive an expression for θ for an ideally banked curve and consider an example related to it.

For **ideal banking**, the net external force equals the horizontal centripetal force in the absence of friction. The components of the normal force N in the horizontal and vertical directions must equal the centripetal force and the weight of the car, respectively. In cases in which forces are not parallel, it is most convenient to consider components along perpendicular axes—in this case, the vertical and horizontal directions.

[\[link\]](#) shows a free body diagram for a car on a frictionless banked curve. If the angle θ is ideal for the speed and radius, then the net external force will equal the necessary centripetal force. The only two external forces acting on the car are its weight \mathbf{w} and the normal force of the road \mathbf{N} . (A frictionless surface can only exert a force perpendicular to the surface—that is, a normal force.) These two forces must add to give a net external force that is horizontal toward the center of curvature and has magnitude mv^2/r .

Because this is the crucial force and it is horizontal, we use a coordinate system with vertical and horizontal axes. Only the normal force has a horizontal component, and so this must equal the centripetal force—that is,

Equation:

$$N \sin \theta = \frac{mv^2}{r}.$$

Because the car does not leave the surface of the road, the net vertical force must be zero, meaning that the vertical components of the two external forces must be equal in magnitude and opposite in direction. From the figure, we see that the vertical component of the normal force is $N \cos \theta$, and the only other vertical force is the car’s weight. These must be equal in magnitude; thus,

Equation:

$$N \cos \theta = mg.$$

Now we can combine the last two equations to eliminate N and get an expression for θ , as desired. Solving the second equation for $N = mg/(\cos \theta)$, and substituting this into the first yields

Equation:

$$mg \frac{\sin \theta}{\cos \theta} = \frac{mv^2}{r}$$

Equation:

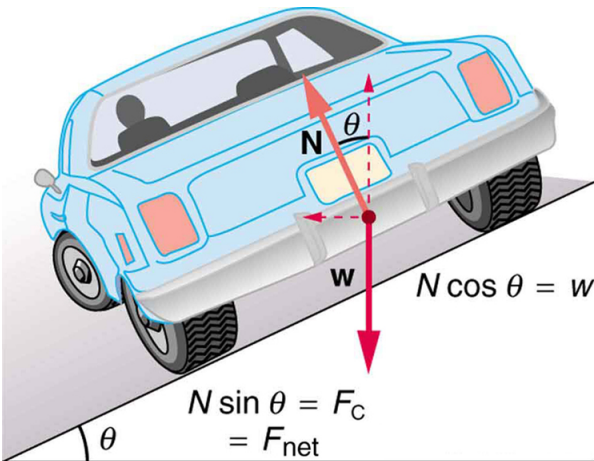
$$\begin{aligned} mg \tan(\theta) &= \frac{mv^2}{r} \\ \tan \theta &= \frac{v^2}{rg} \end{aligned}$$

Taking the inverse tangent gives

Equation:

$$\theta = \tan^{-1} \left(\frac{v^2}{rg} \right) \text{ (ideally banked curve, no friction).}$$

This expression can be understood by considering how θ depends on v and r . A large θ will be obtained for a large v and a small r . That is, roads must be steeply banked for high speeds and sharp curves. Friction helps, because it allows you to take the curve at greater or lower speed than if the curve is frictionless. Note that θ does not depend on the mass of the vehicle.



The car on this banked curve is moving away and turning to the left.

Example:

What Is the Ideal Speed to Take a Steeply Banked Tight Curve?

Curves on some test tracks and race courses, such as the Daytona International Speedway in Florida, are very steeply banked. This banking, with the aid of tire friction and very stable car configurations, allows the curves to be taken at very high speed. To illustrate, calculate the speed at which a 100 m radius curve banked at 65.0° should be driven if the road is frictionless.

Strategy

We first note that all terms in the expression for the ideal angle of a banked curve except for speed are known; thus, we need only rearrange it so that speed appears on the left-hand side and then substitute known quantities.

Solution

Starting with

Equation:

$$\tan \theta = \frac{v^2}{rg}$$

we get

Equation:

$$v = (rg \tan \theta)^{1/2}.$$

Noting that $\tan 65.0^\circ = 2.14$, we obtain

Equation:

$$\begin{aligned} v &= \left[(100 \text{ m})(9.80 \text{ m/s}^2)(2.14) \right]^{1/2} \\ &= 45.8 \text{ m/s.} \end{aligned}$$

Discussion

This is just about 165 km/h, consistent with a very steeply banked and rather sharp curve. Tire friction enables a vehicle to take the curve at significantly higher speeds.

Calculations similar to those in the preceding examples can be performed for a host of interesting situations in which centripetal force is involved—a number of these are presented in this chapter's Problems and Exercises.

Note:

Take-Home Experiment

Ask a friend or relative to swing a golf club or a tennis racquet. Take appropriate measurements to estimate the centripetal acceleration of the end of the club or racquet. You may choose to do this in slow motion.

Note:

PhET Explorations: Gravity and Orbits

Move the sun, earth, moon and space station to see how it affects their gravitational forces and orbital paths. Visualize the sizes and distances between different heavenly bodies, and turn off gravity to see what would happen without it!

Section Summary

- Centripetal force F_c is any force causing uniform circular motion. It is a “center-seeking” force that always points toward the center of rotation. It is perpendicular to linear velocity v and has magnitude

Equation:

$$F_c = ma_c,$$

which can also be expressed as

Equation:

$$\left. \begin{array}{l} F_c = m \frac{v^2}{r} \\ \text{or} \\ F_c = mr\omega^2 \end{array} \right\}$$

Conceptual Questions

Exercise:

Problem:

If you wish to reduce the stress (which is related to centripetal force) on high-speed tires, would you use large- or small-diameter tires? Explain.

Exercise:

Problem:

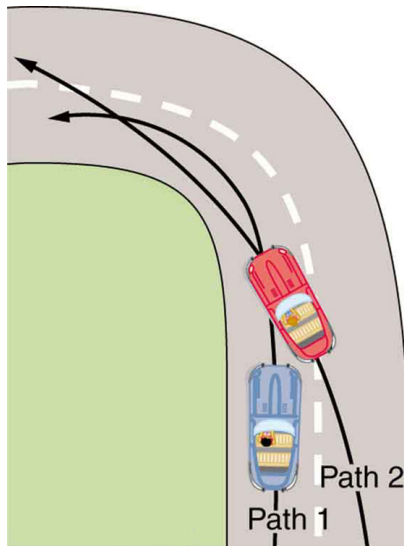
Define centripetal force. Can any type of force (for example, tension, gravitational force, friction, and so on) be a centripetal force? Can any combination of forces be a centripetal force?

Exercise:**Problem:**

If centripetal force is directed toward the center, why do you feel that you are 'thrown' away from the center as a car goes around a curve? Explain.

Exercise:**Problem:**

Race car drivers routinely cut corners as shown in [\[link\]](#). Explain how this allows the curve to be taken at the greatest speed.



Two paths around a race track curve are shown. Race car drivers will take the inside path (called cutting the corner)

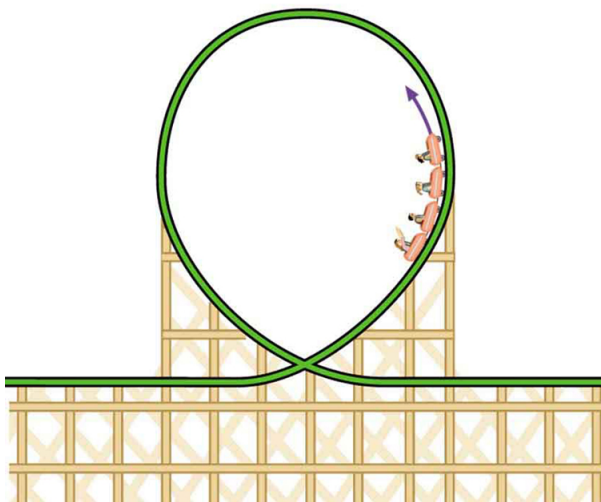
whenever possible
because it allows
them to take the
curve at the highest
speed.

Exercise:

Problem:

A number of amusement parks have rides that make vertical loops like the one shown in [\[link\]](#). For safety, the cars are attached to the rails in such a way that they cannot fall off. If the car goes over the top at just the right speed, gravity alone will supply the centripetal force. What other force acts and what is its direction if:

- (a) The car goes over the top at faster than this speed?
- (b) The car goes over the top at slower than this speed?



Amusement rides with a vertical
loop are an example of a form
of curved motion.

Exercise:**Problem:**

What is the direction of the force exerted by the car on the passenger as the car goes over the top of the amusement ride pictured in [\[link\]](#) under the following circumstances:

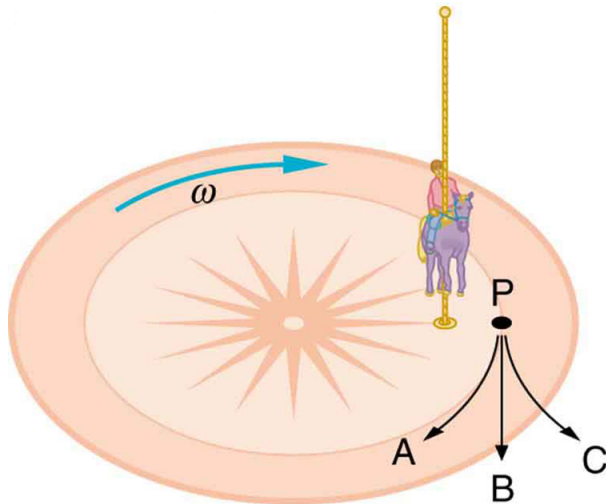
- (a) The car goes over the top at such a speed that the gravitational force is the only force acting?
- (b) The car goes over the top faster than this speed?
- (c) The car goes over the top slower than this speed?

Exercise:**Problem:**

As a skater forms a circle, what force is responsible for making her turn? Use a free body diagram in your answer.

Exercise:**Problem:**

Suppose a child is riding on a merry-go-round at a distance about halfway between its center and edge. She has a lunch box resting on wax paper, so that there is very little friction between it and the merry-go-round. Which path shown in [\[link\]](#) will the lunch box take when she lets go? The lunch box leaves a trail in the dust on the merry-go-round. Is that trail straight, curved to the left, or curved to the right? Explain your answer.



Merry-go-round's rotating
frame of reference

A child riding on a merry-go-round releases her lunch box at point P. This is a view from above the clockwise rotation.

Assuming it slides with negligible friction, will it follow path A, B, or C, as viewed from Earth's frame of reference?

What will be the shape of the path it leaves in the dust on the merry-go-round?

Exercise:

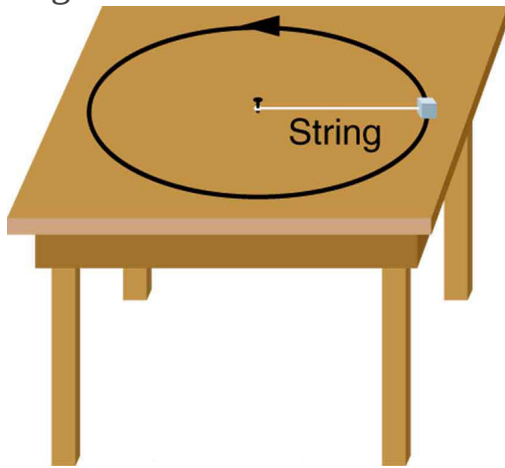
Problem:

Do you feel yourself thrown to either side when you negotiate a curve that is ideally banked for your car's speed? What is the direction of the force exerted on you by the car seat?

Exercise:

Problem:

Suppose a mass is moving in a circular path on a frictionless table as shown in figure. In the Earth's frame of reference, there is no centrifugal force pulling the mass away from the centre of rotation, yet there is a very real force stretching the string attaching the mass to the nail. Using concepts related to centripetal force and Newton's third law, explain what force stretches the string, identifying its physical origin.



A mass attached to a nail on a frictionless table moves in a circular path. The force stretching the string is real and not fictional. What is the physical origin of the force on the string?

Problems Exercise**Exercise:**

Problem:

(a) A 22.0 kg child is riding a playground merry-go-round that is rotating at 40.0 rev/min. What centripetal force must she exert to stay on if she is 1.25 m from its center?

(b) What centripetal force does she need to stay on an amusement park merry-go-round that rotates at 3.00 rev/min if she is 8.00 m from its center?

(c) Compare each force with her weight.

Solution:

a) 483 N

b) 17.4 N

c) 2.24 times her weight, 0.0807 times her weight

Exercise:**Problem:**

Calculate the centripetal force on the end of a 100 m (radius) wind turbine blade that is rotating at 0.5 rev/s. Assume the mass is 4 kg.

Exercise:**Problem:**

What is the ideal banking angle for a gentle turn of 1.20 km radius on a highway with a 105 km/h speed limit (about 65 mi/h), assuming everyone travels at the limit?

Solution:

4.14°

Exercise:

Problem:

What is the ideal speed to take a 100 m radius curve banked at a 20.0° angle?

Exercise:**Problem:**

- (a) What is the radius of a bobsled turn banked at 75.0° and taken at 30.0 m/s, assuming it is ideally banked?
 - (b) Calculate the centripetal acceleration.
 - (c) Does this acceleration seem large to you?
-

Solution:

- a) 24.6 m
- b) 36.6 m/s^2
- c) $a_c = 3.73 g$. This does not seem too large, but it is clear that bobsledders feel a lot of force on them going through sharply banked turns.

Exercise:**Problem:**

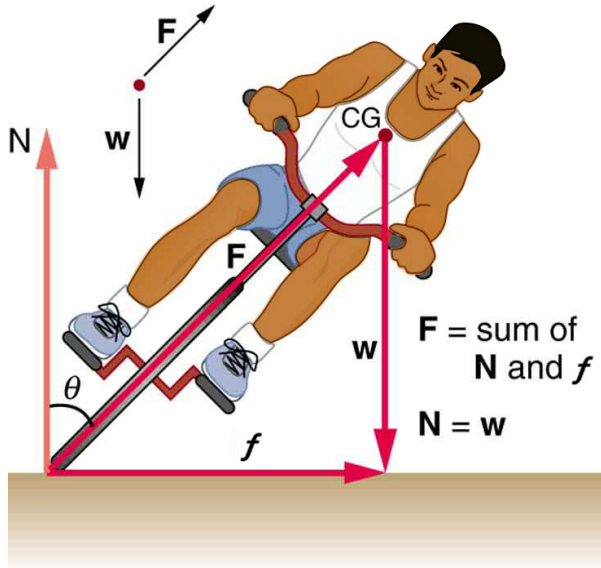
Part of riding a bicycle involves leaning at the correct angle when making a turn, as seen in [\[link\]](#). To be stable, the force exerted by the ground must be on a line going through the center of gravity. The force on the bicycle wheel can be resolved into two perpendicular components—friction parallel to the road (this must supply the centripetal force), and the vertical normal force (which must equal the system's weight).

- (a) Show that θ (as defined in the figure) is related to the speed v and radius of curvature r of the turn in the same way as for an ideally

banked roadway—that is, $\theta = \tan^{-1} v^2 / rg$

(b) Calculate θ for a 12.0 m/s turn of radius 30.0 m (as in a race).

Free-body diagram



A bicyclist negotiating a turn on level ground must lean at the correct angle—the ability to do this becomes instinctive. The force of the ground on the wheel needs to be on a line through the center of gravity. The net external force on the system is the centripetal force. The vertical component of the force on the wheel cancels the weight of the system while its horizontal component must supply the centripetal force.

This process produces a relationship among the angle θ , the speed v , and the radius of curvature r of the turn similar to

that for the ideal banking of
roadways.

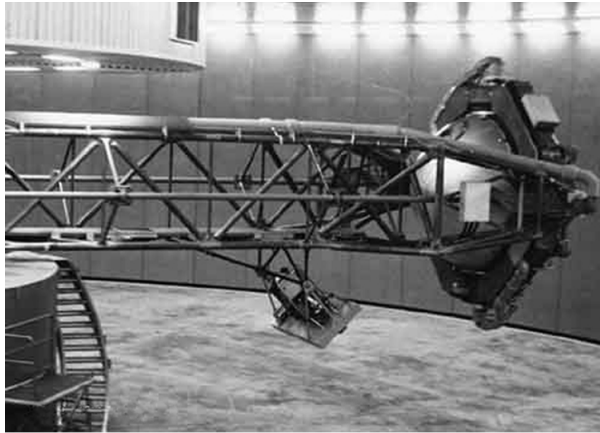
Exercise:

Problem:

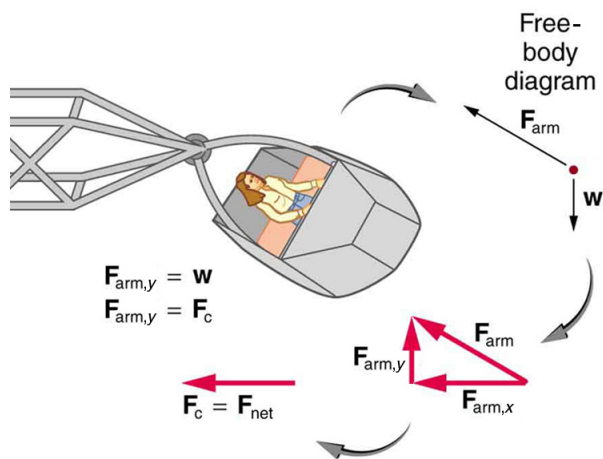
A large centrifuge, like the one shown in [\[link\]](#)(a), is used to expose aspiring astronauts to accelerations similar to those experienced in rocket launches and atmospheric reentries.

(a) At what angular velocity is the centripetal acceleration $10\ g$ if the rider is 15.0 m from the center of rotation?

(b) The rider's cage hangs on a pivot at the end of the arm, allowing it to swing outward during rotation as shown in [\[link\]](#)(b). At what angle θ below the horizontal will the cage hang when the centripetal acceleration is $10\ g$? (Hint: The arm supplies centripetal force and supports the weight of the cage. Draw a free body diagram of the forces to see what the angle θ should be.)



(a) NASA centrifuge and ride



(b)

(a) NASA centrifuge used to subject trainees to accelerations similar to those experienced in rocket launches and reentries.

(credit: NASA) (b) Rider in cage showing how the cage pivots outward during rotation.

This allows the total force exerted on the rider by the cage to be along its axis at all times.

Solution:

a) 2.56 rad/s

b) 5.71°

Exercise:

Problem: Integrated Concepts

If a car takes a banked curve at less than the ideal speed, friction is needed to keep it from sliding toward the inside of the curve (a real problem on icy mountain roads). (a) Calculate the ideal speed to take a 100 m radius curve banked at 15.0° . (b) What is the minimum coefficient of friction needed for a frightened driver to take the same curve at 20.0 km/h?

Solution:

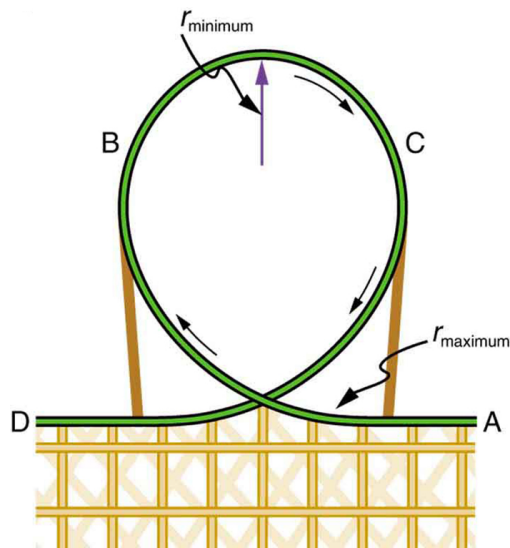
a) 16.2 m/s

b) 0.234

Exercise:

Problem:

Modern roller coasters have vertical loops like the one shown in [\[link\]](#). The radius of curvature is smaller at the top than on the sides so that the downward centripetal acceleration at the top will be greater than the acceleration due to gravity, keeping the passengers pressed firmly into their seats. What is the speed of the roller coaster at the top of the loop if the radius of curvature there is 15.0 m and the downward acceleration of the car is 1.50 g?



Teardrop-shaped loops are used in the latest roller coasters so that the radius of curvature gradually decreases to a minimum at the top. This means that the centripetal acceleration builds from zero to a maximum at the top and gradually decreases again. A circular loop would cause a jolting change in acceleration at entry, a disadvantage discovered long ago in railroad curve design. With a small radius of curvature at the top, the centripetal acceleration can more easily be kept greater than g so that the passengers do not lose contact with their seats nor do they

need seat belts to keep
them in place.

Exercise:

Problem: Unreasonable Results

- (a) Calculate the minimum coefficient of friction needed for a car to negotiate an unbanked 50.0 m radius curve at 30.0 m/s.
 - (b) What is unreasonable about the result?
 - (c) Which premises are unreasonable or inconsistent?
-

Solution:

- a) 1.84
- b) A coefficient of friction this much greater than 1 is unreasonable .
- c) The assumed speed is too great for the tight curve.

Glossary

centripetal force

any net force causing uniform circular motion

ideal banking

the sloping of a curve in a road, where the angle of the slope allows the vehicle to negotiate the curve at a certain speed without the aid of friction between the tires and the road; the net external force on the vehicle equals the horizontal centripetal force in the absence of friction

ideal speed

the maximum safe speed at which a vehicle can turn on a curve without the aid of friction between the tire and the road

ideal angle

the angle at which a car can turn safely on a steep curve, which is in proportion to the ideal speed

banked curve

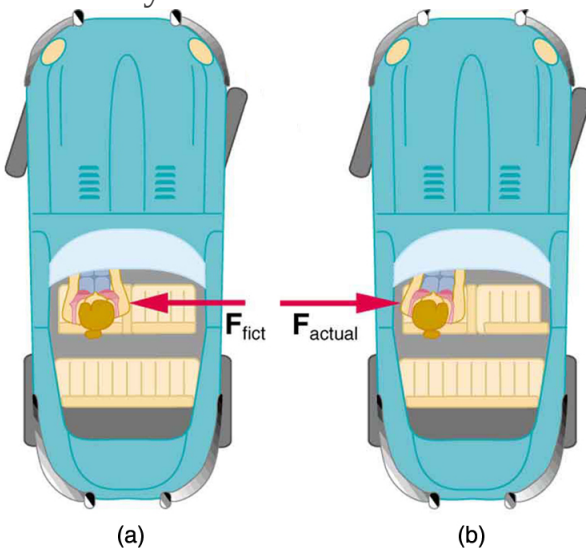
the curve in a road that is sloping in a manner that helps a vehicle negotiate the curve

Fictitious Forces and Non-inertial Frames: The Coriolis Force

- Discuss the inertial frame of reference.
- Discuss the non-inertial frame of reference.
- Describe the effects of the Coriolis force.

What do taking off in a jet airplane, turning a corner in a car, riding a merry-go-round, and the circular motion of a tropical cyclone have in common? Each exhibits fictitious forces—unreal forces that arise from motion and may *seem* real, because the observer's frame of reference is accelerating or rotating.

When taking off in a jet, most people would agree it feels as if you are being pushed back into the seat as the airplane accelerates down the runway. Yet a physicist would say that *you* tend to remain stationary while the *seat* pushes forward on you, and there is no real force backward on you. An even more common experience occurs when you make a tight curve in your car—say, to the right. You feel as if you are thrown (that is, *forced*) toward the left relative to the car. Again, a physicist would say that *you* are going in a straight line but the *car* moves to the right, and there is no real force on you to the left. Recall Newton's first law.

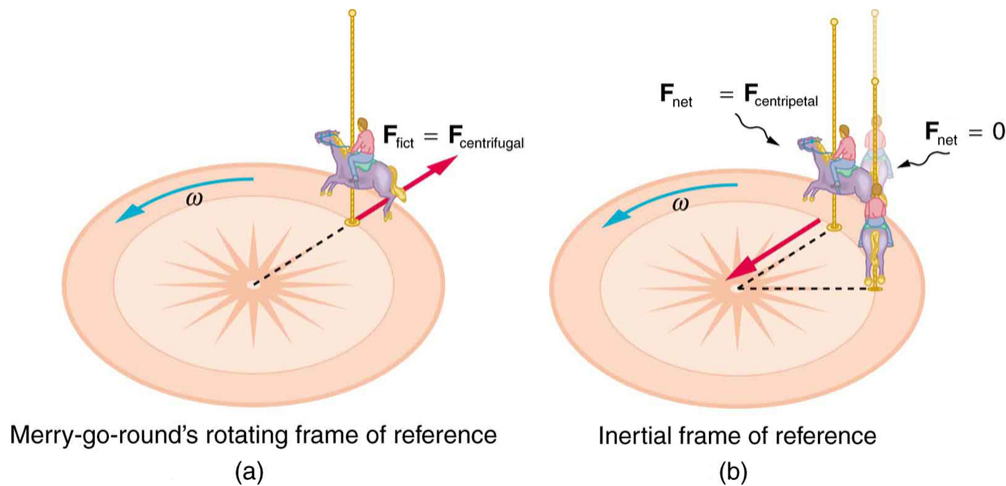


(a) The car driver feels herself forced to the left relative to the car when she makes a right turn. This is a fictitious force arising

from the use of the car as a frame of reference. (b) In the Earth's frame of reference, the driver moves in a straight line, obeying Newton's first law, and the car moves to the right. There is no real force to the left on the driver relative to Earth. There is a real force to the right on the car to make it turn.

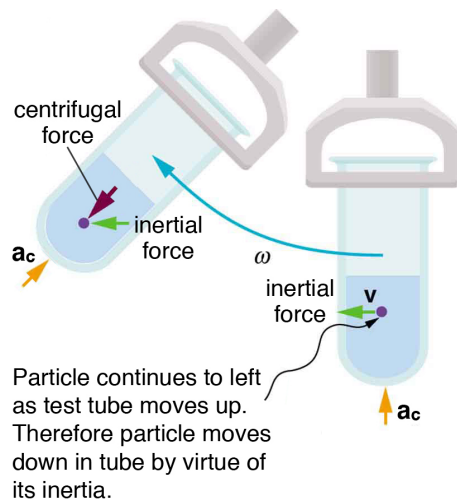
We can reconcile these points of view by examining the frames of reference used. Let us concentrate on people in a car. Passengers instinctively use the car as a frame of reference, while a physicist uses Earth. The physicist chooses Earth because it is very nearly an inertial frame of reference—one in which all forces are real (that is, in which all forces have an identifiable physical origin). In such a frame of reference, Newton's laws of motion take the form given in [Dynamics: Newton's Laws of Motion](#). The car is a **non-inertial frame of reference** because it is accelerated to the side. The force to the left sensed by car passengers is a **fictitious force** having no physical origin. There is nothing real pushing them left—the car, as well as the driver, is actually accelerating to the right.

Let us now take a mental ride on a merry-go-round—specifically, a rapidly rotating playground merry-go-round. You take the merry-go-round to be your frame of reference because you rotate together. In that non-inertial frame, you feel a fictitious force, named **centrifugal force** (not to be confused with centripetal force), trying to throw you off. You must hang on tightly to counteract the centrifugal force. In Earth's frame of reference, there is no force trying to throw you off. Rather you must hang on to make yourself go in a circle because otherwise you would go in a straight line, right off the merry-go-round.



(a) A rider on a merry-go-round feels as if he is being thrown off. This fictitious force is called the centrifugal force—it explains the rider's motion in the rotating frame of reference. (b) In an inertial frame of reference and according to Newton's laws, it is his inertia that carries him off and not a real force (the unshaded rider has $F_{\text{net}} = 0$ and heads in a straight line). A real force, $F_{\text{centripetal}}$, is needed to cause a circular path.

This inertial effect, carrying you away from the center of rotation if there is no centripetal force to cause circular motion, is put to good use in centrifuges (see [\[link\]](#)). A centrifuge spins a sample very rapidly, as mentioned earlier in this chapter. Viewed from the rotating frame of reference, the fictitious centrifugal force throws particles outward, hastening their sedimentation. The greater the angular velocity, the greater the centrifugal force. But what really happens is that the inertia of the particles carries them along a line tangent to the circle while the test tube is forced in a circular path by a centripetal force.

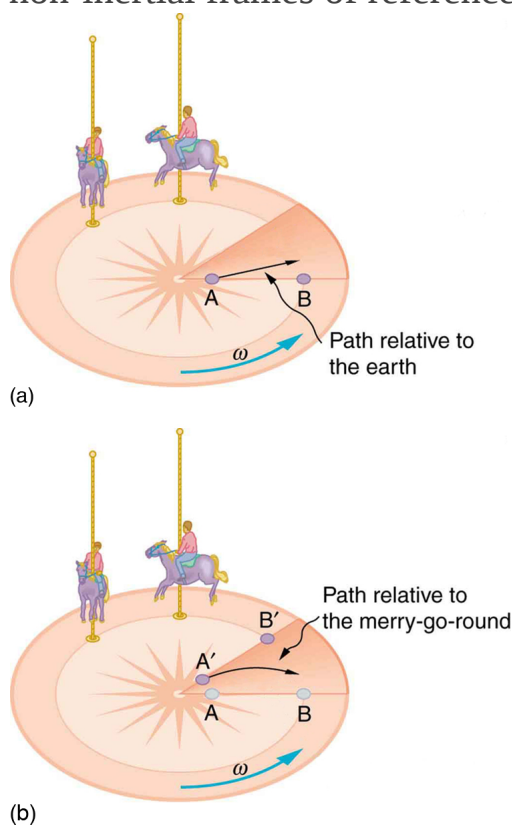


Centrifuges use inertia to perform their task. Particles in the fluid sediment come out because their inertia carries them away from the center of rotation. The large angular velocity of the centrifuge quickens the sedimentation.

Ultimately, the particles will come into contact with the test tube walls, which will then supply the centripetal force needed to make them move in a circle of constant radius.

Let us now consider what happens if something moves in a frame of reference that rotates. For example, what if you slide a ball directly away

from the center of the merry-go-round, as shown in [\[link\]](#)? The ball follows a straight path relative to Earth (assuming negligible friction) and a path curved to the right on the merry-go-round's surface. A person standing next to the merry-go-round sees the ball moving straight and the merry-go-round rotating underneath it. In the merry-go-round's frame of reference, we explain the apparent curve to the right by using a fictitious force, called the **Coriolis force**, that causes the ball to curve to the right. The fictitious Coriolis force can be used by anyone in that frame of reference to explain why objects follow curved paths and allows us to apply Newton's Laws in non-inertial frames of reference.



Looking down on the counterclockwise rotation of a merry-go-round, we see that a ball slid straight toward the edge follows a path curved to the right. The person slides the ball toward point B, starting at

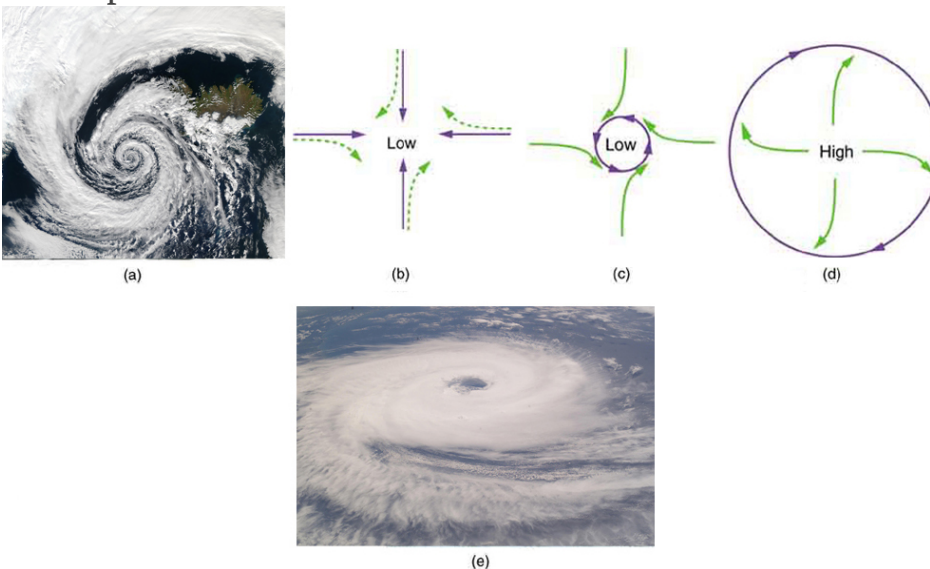
point A. Both points rotate to the shaded positions (A' and B') shown in the time that the ball follows the curved path in the rotating frame and a straight path in Earth's frame.

Up until now, we have considered Earth to be an inertial frame of reference with little or no worry about effects due to its rotation. Yet such effects *do* exist—in the rotation of weather systems, for example. Most consequences of Earth's rotation can be qualitatively understood by analogy with the merry-go-round. Viewed from above the North Pole, Earth rotates counterclockwise, as does the merry-go-round in [\[link\]](#). As on the merry-go-round, any motion in Earth's northern hemisphere experiences a Coriolis force to the right. Just the opposite occurs in the southern hemisphere; there, the force is to the left. Because Earth's angular velocity is small, the Coriolis force is usually negligible, but for large-scale motions, such as wind patterns, it has substantial effects.

The Coriolis force causes hurricanes in the northern hemisphere to rotate in the counterclockwise direction, while the tropical cyclones (what hurricanes are called below the equator) in the southern hemisphere rotate in the clockwise direction. The terms hurricane, typhoon, and tropical storm are regionally-specific names for tropical cyclones, storm systems characterized by low pressure centers, strong winds, and heavy rains. [\[link\]](#) helps show how these rotations take place. Air flows toward any region of low pressure, and tropical cyclones contain particularly low pressures. Thus winds flow toward the center of a tropical cyclone or a low-pressure weather system at the surface. In the northern hemisphere, these inward winds are deflected to the right, as shown in the figure, producing a counterclockwise circulation at the surface for low-pressure zones of any type. Low pressure at the surface is associated with rising air, which also produces cooling and cloud formation, making low-pressure patterns quite

visible from space. Conversely, wind circulation around high-pressure zones is clockwise in the northern hemisphere but is less visible because high pressure is associated with sinking air, producing clear skies.

The rotation of tropical cyclones and the path of a ball on a merry-go-round can just as well be explained by inertia and the rotation of the system underneath. When non-inertial frames are used, fictitious forces, such as the Coriolis force, must be invented to explain the curved path. There is no identifiable physical source for these fictitious forces. In an inertial frame, inertia explains the path, and no force is found to be without an identifiable source. Either view allows us to describe nature, but a view in an inertial frame is the simplest and truest, in the sense that all forces have real origins and explanations.



- (a) The counterclockwise rotation of this northern hemisphere hurricane is a major consequence of the Coriolis force. (credit: NASA) (b) Without the Coriolis force, air would flow straight into a low-pressure zone, such as that found in tropical cyclones. (c) The Coriolis force deflects the winds to the right, producing a counterclockwise rotation. (d) Wind flowing away from a high-pressure zone is also deflected to the right, producing a clockwise rotation. (e) The opposite direction of rotation is produced by

the Coriolis force in the southern hemisphere, leading to tropical cyclones. (credit: NASA)

Section Summary

- Rotating and accelerated frames of reference are non-inertial.
- Fictitious forces, such as the Coriolis force, are needed to explain motion in such frames.

Conceptual Questions

Exercise:

Problem:

When a toilet is flushed or a sink is drained, the water (and other material) begins to rotate about the drain on the way down. Assuming no initial rotation and a flow initially directly straight toward the drain, explain what causes the rotation and which direction it has in the northern hemisphere. (Note that this is a small effect and in most toilets the rotation is caused by directional water jets.) Would the direction of rotation reverse if water were forced up the drain?

Exercise:

Problem:

Is there a real force that throws water from clothes during the spin cycle of a washing machine? Explain how the water is removed.

Exercise:

Problem:

In one amusement park ride, riders enter a large vertical barrel and stand against the wall on its horizontal floor. The barrel is spun up and the floor drops away. Riders feel as if they are pinned to the wall by a force something like the gravitational force. This is a fictitious force sensed and used by the riders to explain events in the rotating frame of reference of the barrel. Explain in an inertial frame of reference (Earth is nearly one) what pins the riders to the wall, and identify all of the real forces acting on them.

Exercise:**Problem:**

Action at a distance, such as is the case for gravity, was once thought to be illogical and therefore untrue. What is the ultimate determinant of the truth in physics, and why was this action ultimately accepted?

Exercise:**Problem:**

Two friends are having a conversation. Anna says a satellite in orbit is in freefall because the satellite keeps falling toward Earth. Tom says a satellite in orbit is not in freefall because the acceleration due to gravity is not 9.80 m/s^2 . Who do you agree with and why?

Exercise:**Problem:**

A non-rotating frame of reference placed at the center of the Sun is very nearly an inertial one. Why is it not exactly an inertial frame?

Glossary

fictitious force

a force having no physical origin

centrifugal force

a fictitious force that tends to throw an object off when the object is rotating in a non-inertial frame of reference

Coriolis force

the fictitious force causing the apparent deflection of moving objects when viewed in a rotating frame of reference

non-inertial frame of reference

an accelerated frame of reference

Newton's Universal Law of Gravitation

- Explain Earth's gravitational force.
- Describe the gravitational effect of the Moon on Earth.
- Discuss weightlessness in space.
- Examine the Cavendish experiment

What do aching feet, a falling apple, and the orbit of the Moon have in common? Each is caused by the gravitational force. Our feet are strained by supporting our weight—the force of Earth's gravity on us. An apple falls from a tree because of the same force acting a few meters above Earth's surface. And the Moon orbits Earth because gravity is able to supply the necessary centripetal force at a distance of hundreds of millions of meters. In fact, the same force causes planets to orbit the Sun, stars to orbit the center of the galaxy, and galaxies to cluster together. Gravity is another example of underlying simplicity in nature. It is the weakest of the four basic forces found in nature, and in some ways the least understood. It is a force that acts at a distance, without physical contact, and is expressed by a formula that is valid everywhere in the universe, for masses and distances that vary from the tiny to the immense.

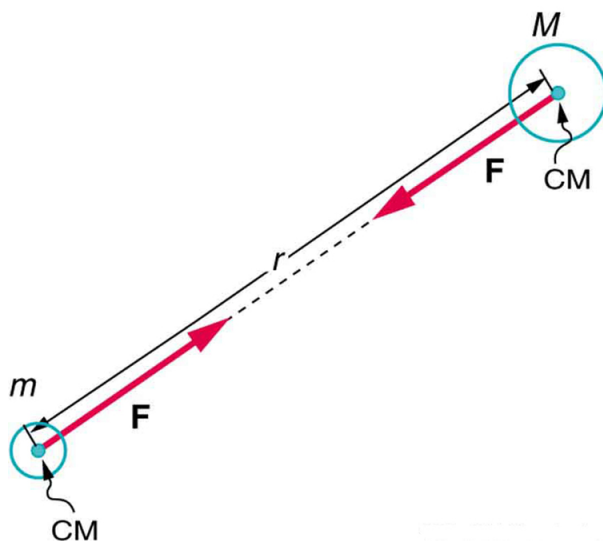
Sir Isaac Newton was the first scientist to precisely define the gravitational force, and to show that it could explain both falling bodies and astronomical motions. See [\[link\]](#). But Newton was not the first to suspect that the same force caused both our weight and the motion of planets. His forerunner Galileo Galilei had contended that falling bodies and planetary motions had the same cause. Some of Newton's contemporaries, such as Robert Hooke, Christopher Wren, and Edmund Halley, had also made some progress toward understanding gravitation. But Newton was the first to propose an exact mathematical form and to use that form to show that the motion of heavenly bodies should be conic sections—circles, ellipses, parabolas, and hyperbolas. This theoretical prediction was a major triumph—it had been known for some time that moons, planets, and comets follow such paths, but no one had been able to propose a mechanism that caused them to follow these paths and not others.



According to early accounts, Newton was inspired to make the connection between falling bodies and astronomical motions when he saw an apple fall from a tree and realized that if the gravitational force could extend above the ground to a tree, it might also reach the Sun. The inspiration of Newton's apple is a part of worldwide folklore and may even be based in fact. Great importance is attached to it because Newton's

universal law of
gravitation and his
laws of motion
answered very old
questions about nature
and gave tremendous
support to the notion
of underlying
simplicity and unity in
nature. Scientists still
expect underlying
simplicity to emerge
from their ongoing
inquiries into nature.

The gravitational force is relatively simple. It is always attractive, and it depends only on the masses involved and the distance between them. Stated in modern language, **Newton's universal law of gravitation** states that every particle in the universe attracts every other particle with a force along a line joining them. The force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.



Gravitational attraction is along a line joining the centers of mass of these two bodies. The magnitude of the force is the same on each, consistent with Newton's third law.

Note:

Misconception Alert

The magnitude of the force on each object (one has larger mass than the other) is the same, consistent with Newton's third law.

The bodies we are dealing with tend to be large. To simplify the situation we assume that the body acts as if its entire mass is concentrated at one specific point called the **center of mass** (CM), which will be further explored in [Linear Momentum and Collisions](#). For two bodies having masses m and M with a distance r between their centers of mass, the equation for Newton's universal law of gravitation is

Equation:

$$F = G \frac{mM}{r^2},$$

where F is the magnitude of the gravitational force and G is a proportionality factor called the **gravitational constant**. G is a universal gravitational constant—that is, it is thought to be the same everywhere in the universe. It has been measured experimentally to be

Equation:

$$G = 6.674 \times 10^{-11} \frac{\text{N} \cdot \text{m}^2}{\text{kg}^2}$$

in SI units. Note that the units of G are such that a force in newtons is obtained from $F = G \frac{mM}{r^2}$, when considering masses in kilograms and distance in meters. For example, two 1.000 kg masses separated by 1.000 m will experience a gravitational attraction of 6.674×10^{-11} N. This is an extraordinarily small force. The small magnitude of the gravitational force is consistent with everyday experience. We are unaware that even large objects like mountains exert gravitational forces on us. In fact, our body weight is the force of attraction of the *entire Earth* on us with a mass of 6×10^{24} kg.

Recall that the acceleration due to gravity g is about 9.80 m/s^2 on Earth. We can now determine why this is so. The weight of an object mg is the gravitational force between it and Earth. Substituting mg for F in Newton's universal law of gravitation gives

Equation:

$$mg = G \frac{mM}{r^2},$$

where m is the mass of the object, M is the mass of Earth, and r is the distance to the center of Earth (the distance between the centers of mass of the object and Earth). See [\[link\]](#). The mass m of the object cancels, leaving an equation for g :

Equation:

$$g = G \frac{M}{r^2}.$$

Substituting known values for Earth's mass and radius (to three significant figures),

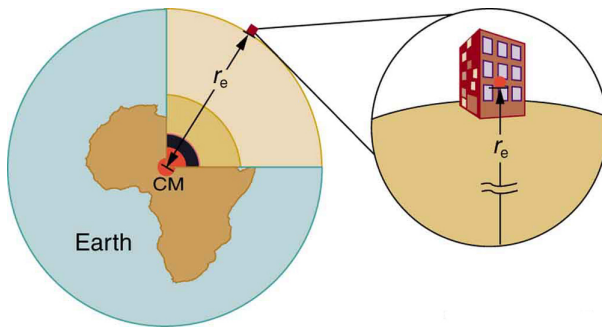
Equation:

$$g = \left(6.67 \times 10^{-11} \frac{\text{N} \cdot \text{m}^2}{\text{kg}^2} \right) \times \frac{5.98 \times 10^{24} \text{ kg}}{(6.38 \times 10^6 \text{ m})^2},$$

and we obtain a value for the acceleration of a falling body:

Equation:

$$g = 9.80 \text{ m/s}^2.$$



The distance between the centers of mass of Earth and an object on its surface is very nearly the same as the radius of Earth, because Earth is so much larger than the object.

This is the expected value *and is independent of the body's mass*. Newton's law of gravitation takes Galileo's observation that all masses fall with the same acceleration a step further, explaining the observation in terms of a force that causes objects to fall—in fact, in terms of a universally existing force of attraction between masses.

Note:

Take-Home Experiment

Take a marble, a ball, and a spoon and drop them from the same height. Do they hit the floor at the same time? If you drop a piece of paper as well, does it behave like the other objects? Explain your observations.

Note:**Making Connections**

Attempts are still being made to understand the gravitational force. As we shall see in [Particle Physics](#), modern physics is exploring the connections of gravity to other forces, space, and time. General relativity alters our view of gravitation, leading us to think of gravitation as bending space and time.

In the following example, we make a comparison similar to one made by Newton himself. He noted that if the gravitational force caused the Moon to orbit Earth, then the acceleration due to gravity should equal the centripetal acceleration of the Moon in its orbit. Newton found that the two accelerations agreed “pretty nearly.”

Example:**Earth’s Gravitational Force Is the Centripetal Force Making the Moon Move in a Curved Path**

- (a) Find the acceleration due to Earth’s gravity at the distance of the Moon.
- (b) Calculate the centripetal acceleration needed to keep the Moon in its orbit (assuming a circular orbit about a fixed Earth), and compare it with the value of the acceleration due to Earth’s gravity that you have just found.

Strategy for (a)

This calculation is the same as the one finding the acceleration due to gravity at Earth’s surface, except that r is the distance from the center of Earth to the center of the Moon. The radius of the Moon’s nearly circular orbit is 3.84×10^8 m.

Solution for (a)

Substituting known values into the expression for g found above, remembering that M is the mass of Earth not the Moon, yields

Equation:

$$\begin{aligned}
 g &= G \frac{M}{r^2} = \left(6.67 \times 10^{-11} \frac{\text{N} \cdot \text{m}^2}{\text{kg}^2} \right) \times \frac{5.98 \times 10^{24} \text{ kg}}{(3.84 \times 10^8 \text{ m})^2} \\
 &= 2.70 \times 10^{-3} \text{ m/s}^2
 \end{aligned}$$

Strategy for (b)

Centripetal acceleration can be calculated using either form of **Equation:**

$$\left. \begin{aligned} a_c &= \frac{v^2}{r} \\ a_c &= r\omega^2 \end{aligned} \right\}$$

We choose to use the second form:

Equation:

$$a_c = r\omega^2,$$

where ω is the angular velocity of the Moon about Earth.

Solution for (b)

Given that the period (the time it takes to make one complete rotation) of the Moon's orbit is 27.3 days, (d) and using

Equation:

$$1 \text{ d} \times 24 \frac{\text{hr}}{\text{d}} \times 60 \frac{\text{min}}{\text{hr}} \times 60 \frac{\text{s}}{\text{min}} = 86,400 \text{ s}$$

we see that

Equation:

$$\omega = \frac{\Delta\theta}{\Delta t} = \frac{2\pi \text{ rad}}{(27.3 \text{ d})(86,400 \text{ s/d})} = 2.66 \times 10^{-6} \frac{\text{rad}}{\text{s}}.$$

The centripetal acceleration is

Equation:

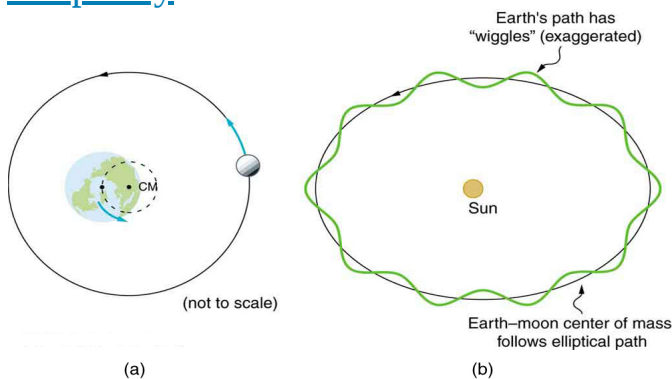
$$\begin{aligned}
 a_c &= r\omega^2 = (3.84 \times 10^8 \text{ m})(2.66 \times 10^{-6} \text{ rad/s})^2 \\
 &= 2.72 \times 10^{-3} \text{ m/s}^2
 \end{aligned}$$

The direction of the acceleration is toward the center of the Earth.

Discussion

The centripetal acceleration of the Moon found in (b) differs by less than 1% from the acceleration due to Earth's gravity found in (a). This agreement is approximate because the Moon's orbit is slightly elliptical, and Earth is not stationary (rather the Earth-Moon system rotates about its center of mass, which is located some 1700 km below Earth's surface). The clear implication is that Earth's gravitational force causes the Moon to orbit Earth.

Why does Earth not remain stationary as the Moon orbits it? This is because, as expected from Newton's third law, if Earth exerts a force on the Moon, then the Moon should exert an equal and opposite force on Earth (see [\[link\]](#)). We do not sense the Moon's effect on Earth's motion, because the Moon's gravity moves our bodies right along with Earth but there are other signs on Earth that clearly show the effect of the Moon's gravitational force as discussed in [Satellites and Kepler's Laws: An Argument for Simplicity](#).

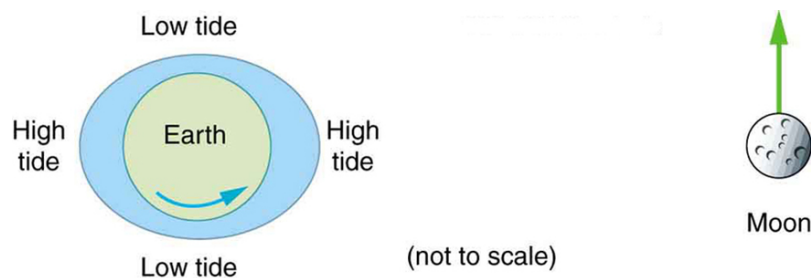


(a) Earth and the Moon rotate approximately once a month around their common center of mass. (b) Their center of mass orbits the Sun in an elliptical orbit, but Earth's path around the Sun has "wiggles" in it. Similar wiggles in the paths of stars have been observed and are

considered direct evidence of planets orbiting those stars. This is important because the planets' reflected light is often too dim to be observed.

Tides

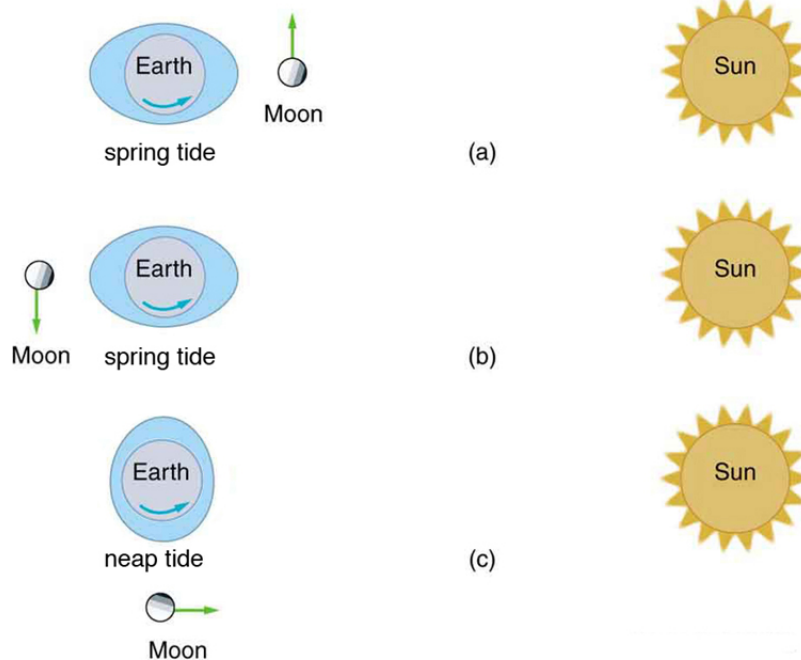
Ocean tides are one very observable result of the Moon's gravity acting on Earth. [\[link\]](#) is a simplified drawing of the Moon's position relative to the tides. Because water easily flows on Earth's surface, a high tide is created on the side of Earth nearest to the Moon, where the Moon's gravitational pull is strongest. Why is there also a high tide on the opposite side of Earth? The answer is that Earth is pulled toward the Moon more than the water on the far side, because Earth is closer to the Moon. So the water on the side of Earth closest to the Moon is pulled away from Earth, and Earth is pulled away from water on the far side. As Earth rotates, the tidal bulge (an effect of the tidal forces between an orbiting natural satellite and the primary planet that it orbits) keeps its orientation with the Moon. Thus there are two tides per day (the actual tidal period is about 12 hours and 25.2 minutes), because the Moon moves in its orbit each day as well).



The Moon causes ocean tides by attracting the water on the near side more than Earth, and by attracting Earth more than the water on the far side. The distances and sizes are

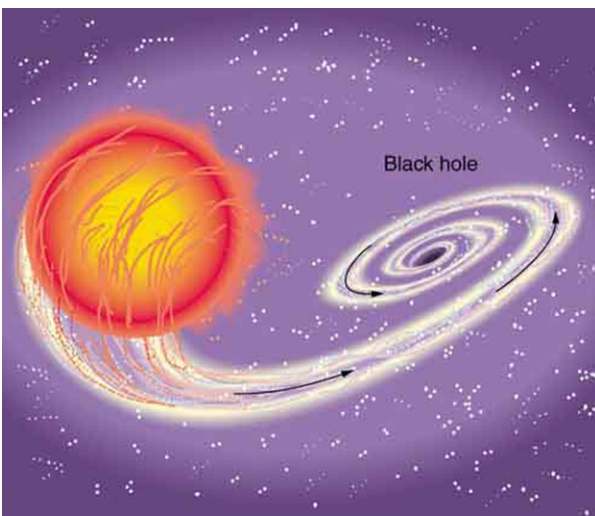
not to scale. For this simplified representation of the Earth-Moon system, there are two high and two low tides per day at any location, because Earth rotates under the tidal bulge.

The Sun also affects tides, although it has about half the effect of the Moon. However, the largest tides, called spring tides, occur when Earth, the Moon, and the Sun are aligned. The smallest tides, called neap tides, occur when the Sun is at a 90° angle to the Earth-Moon alignment.



(a, b) Spring tides: The highest tides occur when Earth, the Moon, and the Sun are aligned. (c) Neap tide: The lowest tides occur when the Sun lies at 90° to the Earth-Moon alignment. Note that this figure is not drawn to scale.

Tides are not unique to Earth but occur in many astronomical systems. The most extreme tides occur where the gravitational force is the strongest and varies most rapidly, such as near black holes (see [\[link\]](#)). A few likely candidates for black holes have been observed in our galaxy. These have masses greater than the Sun but have diameters only a few kilometers across. The tidal forces near them are so great that they can actually tear matter from a companion star.



A black hole is an object with such strong gravity that not even light can escape it. This black hole was created by the supernova of one star in a two-star system. The tidal forces created by the black hole are so great that it tears matter from the companion star. This matter is compressed and heated as it is sucked into the black hole, creating light and X-rays observable from Earth.

”Weightlessness” and Microgravity

In contrast to the tremendous gravitational force near black holes is the apparent gravitational field experienced by astronauts orbiting Earth. What is the effect of “weightlessness” upon an astronaut who is in orbit for months? Or what about the effect of weightlessness upon plant growth? Weightlessness doesn’t mean that an astronaut is not being acted upon by the gravitational force. There is no “zero gravity” in an astronaut’s orbit. The term just means that the astronaut is in free-fall, accelerating with the acceleration due to gravity. If an elevator cable breaks, the passengers inside will be in free fall and will experience weightlessness. You can experience short periods of weightlessness in some rides in amusement parks.



Astronauts experiencing weightlessness on board the International Space Station. (credit: NASA)

Microgravity refers to an environment in which the apparent net acceleration of a body is small compared with that produced by Earth at its surface. Many interesting biology and physics topics have been studied over the past three decades in the presence of microgravity. Of immediate concern is the effect on astronauts of extended times in outer space, such as at the International Space Station. Researchers have observed that muscles will atrophy (waste away) in this environment. There is also a corresponding loss of bone mass. Study continues on cardiovascular adaptation to space flight. On Earth, blood pressure is usually higher in the feet than in the head, because the higher column of blood exerts a downward force on it, due to gravity. When standing, 70% of your blood is below the level of the heart, while in a horizontal position, just the opposite occurs. What difference does the absence of this pressure differential have upon the heart?

Some findings in human physiology in space can be clinically important to the management of diseases back on Earth. On a somewhat negative note, spaceflight is known to affect the human immune system, possibly making the crew members more vulnerable to infectious diseases. Experiments flown in space also have shown that some bacteria grow faster in microgravity than they do on Earth. However, on a positive note, studies indicate that microbial antibiotic production can increase by a factor of two in space-grown cultures. One hopes to be able to understand these mechanisms so that similar successes can be achieved on the ground. In another area of physics space research, inorganic crystals and protein crystals have been grown in outer space that have much higher quality than any grown on Earth, so crystallography studies on their structure can yield much better results.

Plants have evolved with the stimulus of gravity and with gravity sensors. Roots grow downward and shoots grow upward. Plants might be able to provide a life support system for long duration space missions by regenerating the atmosphere, purifying water, and producing food. Some studies have indicated that plant growth and development are not affected by gravity, but there is still uncertainty about structural changes in plants grown in a microgravity environment.

The Cavendish Experiment: Then and Now

As previously noted, the universal gravitational constant G is determined experimentally. This definition was first done accurately by Henry Cavendish (1731–1810), an English scientist, in 1798, more than 100 years after Newton published his universal law of gravitation. The measurement of G is very basic and important because it determines the strength of one of the four forces in nature. Cavendish's experiment was very difficult because he measured the tiny gravitational attraction between two ordinary-sized masses (tens of kilograms at most), using apparatus like that in [\[link\]](#). Remarkably, his value for G differs by less than 1% from the best modern value.

One important consequence of knowing G was that an accurate value for Earth's mass could finally be obtained. This was done by measuring the acceleration due to gravity as accurately as possible and then calculating the mass of Earth M from the relationship Newton's universal law of gravitation gives

Equation:

$$mg = G \frac{mM}{r^2},$$

where m is the mass of the object, M is the mass of Earth, and r is the distance to the center of Earth (the distance between the centers of mass of the object and Earth). See [\[link\]](#). The mass m of the object cancels, leaving an equation for g :

Equation:

$$g = G \frac{M}{r^2}.$$

Rearranging to solve for M yields

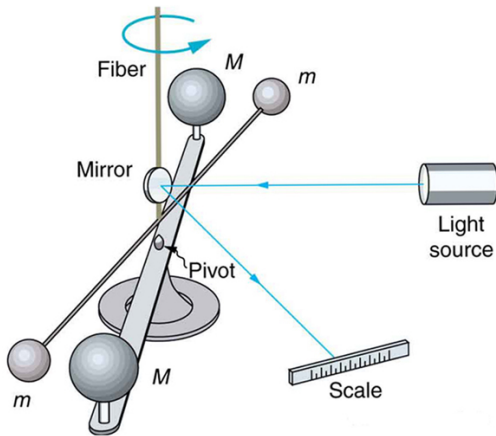
Equation:

$$M = \frac{gr^2}{G}.$$

So M can be calculated because all quantities on the right, including the radius of Earth r , are known from direct measurements. We shall see in [Satellites and Kepler's Laws: An Argument for Simplicity](#) that knowing G also allows for the determination of astronomical masses. Interestingly, of all the fundamental constants in physics, G is by far the least well determined.

The Cavendish experiment is also used to explore other aspects of gravity. One of the most interesting questions is whether the gravitational force depends on substance as well as mass—for example, whether one kilogram of lead exerts the same gravitational pull as one kilogram of water. A Hungarian scientist named Roland von Eötvös pioneered this inquiry early in the 20th century. He found, with an accuracy of five parts per billion, that the gravitational force does not depend on the substance. Such experiments continue today, and have improved upon Eötvös' measurements.

Cavendish-type experiments such as those of Eric Adelberger and others at the University of Washington, have also put severe limits on the possibility of a fifth force and have verified a major prediction of general relativity—that gravitational energy contributes to rest mass. Ongoing measurements there use a torsion balance and a parallel plate (not spheres, as Cavendish used) to examine how Newton's law of gravitation works over sub-millimeter distances. On this small-scale, do gravitational effects depart from the inverse square law? So far, no deviation has been observed.



Cavendish used an apparatus like this to measure the gravitational attraction between the two suspended spheres (m) and the two on the stand (M) by observing the amount of torsion (twisting) created in the fiber. Distance between the masses can be varied to check the dependence of the force on distance. Modern experiments of this type continue to explore gravity.

Section Summary

- Newton's universal law of gravitation: Every particle in the universe attracts every other particle with a force along a line joining them. The force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. In equation form, this is

Equation:

$$F = G \frac{mM}{r^2},$$

where F is the magnitude of the gravitational force. G is the gravitational constant, given by $G = 6.674 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$.

- Newton's law of gravitation applies universally.

Conceptual Questions**Exercise:****Problem:**

Action at a distance, such as is the case for gravity, was once thought to be illogical and therefore untrue. What is the ultimate determinant of the truth in physics, and why was this action ultimately accepted?

Exercise:**Problem:**

Two friends are having a conversation. Anna says a satellite in orbit is in freefall because the satellite keeps falling toward Earth. Tom says a satellite in orbit is not in freefall because the acceleration due to gravity is not 9.80 m/s^2 . Who do you agree with and why?

Exercise:**Problem:**

Draw a free body diagram for a satellite in an elliptical orbit showing why its speed increases as it approaches its parent body and decreases as it moves away.

Exercise:

Problem:

Newton's laws of motion and gravity were among the first to convincingly demonstrate the underlying simplicity and unity in nature. Many other examples have since been discovered, and we now expect to find such underlying order in complex situations. Is there proof that such order will always be found in new explorations?

Problem Exercises**Exercise:****Problem:**

(a) Calculate Earth's mass given the acceleration due to gravity at the North Pole is 9.830 m/s^2 and the radius of the Earth is 6371 km from center to pole.

(b) Compare this with the accepted value of $5.979 \times 10^{24} \text{ kg}$.

Solution:

a) $5.979 \times 10^{24} \text{ kg}$

b) This is identical to the best value to three significant figures.

Exercise:**Problem:**

(a) Calculate the magnitude of the acceleration due to gravity on the surface of Earth due to the Moon.

(b) Calculate the magnitude of the acceleration due to gravity at Earth due to the Sun.

(c) Take the ratio of the Moon's acceleration to the Sun's and comment on why the tides are predominantly due to the Moon in spite of this

number.

Exercise:

Problem:

- (a) What is the acceleration due to gravity on the surface of the Moon?
 - (b) On the surface of Mars? The mass of Mars is 6.418×10^{23} kg and its radius is 3.38×10^6 m.
-

Solution:

- a) 1.62 m/s^2
- b) 3.75 m/s^2

Exercise:

Problem:

- (a) Calculate the acceleration due to gravity on the surface of the Sun.
- (b) By what factor would your weight increase if you could stand on the Sun? (Never mind that you cannot.)

Exercise:

Problem:

The Moon and Earth rotate about their common center of mass, which is located about 4700 km from the center of Earth. (This is 1690 km below the surface.)

- (a) Calculate the magnitude of the acceleration due to the Moon's gravity at that point.
- (b) Calculate the magnitude of the centripetal acceleration of the center of Earth as it rotates about that point once each lunar month (about 27.3 d) and compare it with the acceleration found in part (a).

Comment on whether or not they are equal and why they should or should not be.

Solution:

a) $3.42 \times 10^{-5} \text{ m/s}^2$

b) $3.34 \times 10^{-5} \text{ m/s}^2$

The values are nearly identical. One would expect the gravitational force to be the same as the centripetal force at the core of the system.

Exercise:

Problem: Solve part (b) of [\[link\]](#) using $a_c = v^2/r$.

Exercise:

Problem:

Astrology, that unlikely and vague pseudoscience, makes much of the position of the planets at the moment of one's birth. The only known force a planet exerts on Earth is gravitational.

(a) Calculate the magnitude of the gravitational force exerted on a 4.20 kg baby by a 100 kg father 0.200 m away at birth (he is assisting, so he is close to the child).

(b) Calculate the magnitude of the force on the baby due to Jupiter if it is at its closest distance to Earth, some $6.29 \times 10^{11} \text{ m}$ away. How does the force of Jupiter on the baby compare to the force of the father on the baby? Other objects in the room and the hospital building also exert similar gravitational forces. (Of course, there could be an unknown force acting, but scientists first need to be convinced that there is even an effect, much less that an unknown force causes it.)

Solution:

a) $7.01 \times 10^{-7} \text{ N}$

b) $1.35 \times 10^{-6} \text{ N}$, 0.521

Exercise:

Problem:

The existence of the dwarf planet Pluto was proposed based on irregularities in Neptune's orbit. Pluto was subsequently discovered near its predicted position. But it now appears that the discovery was fortuitous, because Pluto is small and the irregularities in Neptune's orbit were not well known. To illustrate that Pluto has a minor effect on the orbit of Neptune compared with the closest planet to Neptune:

(a) Calculate the acceleration due to gravity at Neptune due to Pluto when they are $4.50 \times 10^{12} \text{ m}$ apart, as they are at present. The mass of Pluto is $1.4 \times 10^{22} \text{ kg}$.

(b) Calculate the acceleration due to gravity at Neptune due to Uranus, presently about $2.50 \times 10^{12} \text{ m}$ apart, and compare it with that due to Pluto. The mass of Uranus is $8.62 \times 10^{25} \text{ kg}$.

Exercise:

Problem:

(a) The Sun orbits the Milky Way galaxy once each $2.60 \times 10^8 \text{ y}$, with a roughly circular orbit averaging 3.00×10^4 light years in radius. (A light year is the distance traveled by light in 1 y.) Calculate the centripetal acceleration of the Sun in its galactic orbit. Does your result support the contention that a nearly inertial frame of reference can be located at the Sun?

(b) Calculate the average speed of the Sun in its galactic orbit. Does the answer surprise you?

Solution:

a) $1.66 \times 10^{-10} \text{ m/s}^2$

b) $2.17 \times 10^5 \text{ m/s}$

Exercise:

Problem: Unreasonable Result

A mountain 10.0 km from a person exerts a gravitational force on him equal to 2.00% of his weight.

- (a) Calculate the mass of the mountain.
 - (b) Compare the mountain's mass with that of Earth.
 - (c) What is unreasonable about these results?
 - (d) Which premises are unreasonable or inconsistent? (Note that accurate gravitational measurements can easily detect the effect of nearby mountains and variations in local geology.)
-

Solution:

a) $2.937 \times 10^{17} \text{ kg}$

b) 4.91×10^{-8}

of the Earth's mass.

c) The mass of the mountain and its fraction of the Earth's mass are too great.

d) The gravitational force assumed to be exerted by the mountain is too great.

Glossary

gravitational constant, G

a proportionality factor used in the equation for Newton's universal law of gravitation; it is a universal constant—that is, it is thought to be the same everywhere in the universe

center of mass

the point where the entire mass of an object can be thought to be concentrated

microgravity

an environment in which the apparent net acceleration of a body is small compared with that produced by Earth at its surface

Newton's universal law of gravitation

every particle in the universe attracts every other particle with a force along a line joining them; the force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them

Satellites and Kepler's Laws: An Argument for Simplicity

- State Kepler's laws of planetary motion.
- Derive the third Kepler's law for circular orbits.
- Discuss the Ptolemaic model of the universe.

Examples of gravitational orbits abound. Hundreds of artificial satellites orbit Earth together with thousands of pieces of debris. The Moon's orbit about Earth has intrigued humans from time immemorial. The orbits of planets, asteroids, meteors, and comets about the Sun are no less interesting. If we look further, we see almost unimaginable numbers of stars, galaxies, and other celestial objects orbiting one another and interacting through gravity.

All these motions are governed by gravitational force, and it is possible to describe them to various degrees of precision. Precise descriptions of complex systems must be made with large computers. However, we can describe an important class of orbits without the use of computers, and we shall find it instructive to study them. These orbits have the following characteristics:

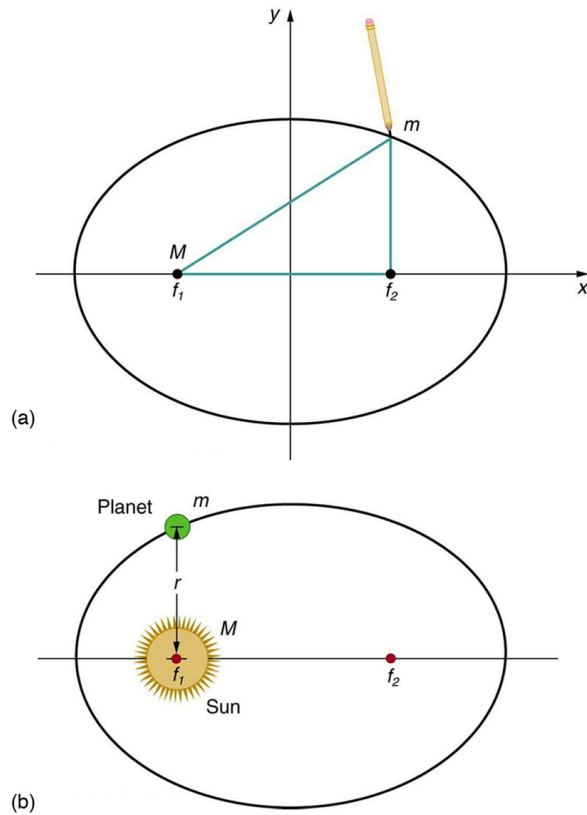
1. *A small mass m orbits a much larger mass M .* This allows us to view the motion as if M were stationary—in fact, as if from an inertial frame of reference placed on M —without significant error. Mass m is the satellite of M , if the orbit is gravitationally bound.
2. *The system is isolated from other masses.* This allows us to neglect any small effects due to outside masses.

The conditions are satisfied, to good approximation, by Earth's satellites (including the Moon), by objects orbiting the Sun, and by the satellites of other planets. Historically, planets were studied first, and there is a classical set of three laws, called Kepler's laws of planetary motion, that describe the orbits of all bodies satisfying the two previous conditions (not just planets in our solar system). These descriptive laws are named for the German astronomer Johannes Kepler (1571–1630), who devised them after careful study (over some 20 years) of a large amount of meticulously recorded observations of planetary motion done by Tycho Brahe (1546–1601). Such careful collection and detailed recording of methods and data are hallmarks of good science. Data constitute the evidence from which new interpretations and meanings can be constructed.

Kepler's Laws of Planetary Motion

Kepler's First Law

The orbit of each planet about the Sun is an ellipse with the Sun at one focus.



(a) An ellipse is a closed curve such that the sum of the distances from a point on the curve to the two foci (f_1 and f_2) is a constant. You can draw an ellipse as shown by putting a pin at each focus, and then placing a string around a pencil and the pins and tracing a line on paper. A circle is a special case of an ellipse in which the

two foci coincide (thus any point on the circle is the same distance from the center). (b)

For any closed gravitational orbit, m follows an elliptical path with M at one focus.

Kepler's first law states this fact for planets orbiting the Sun.

Kepler's Second Law

Each planet moves so that an imaginary line drawn from the Sun to the planet sweeps out equal areas in equal times (see [\[link\]](#)).

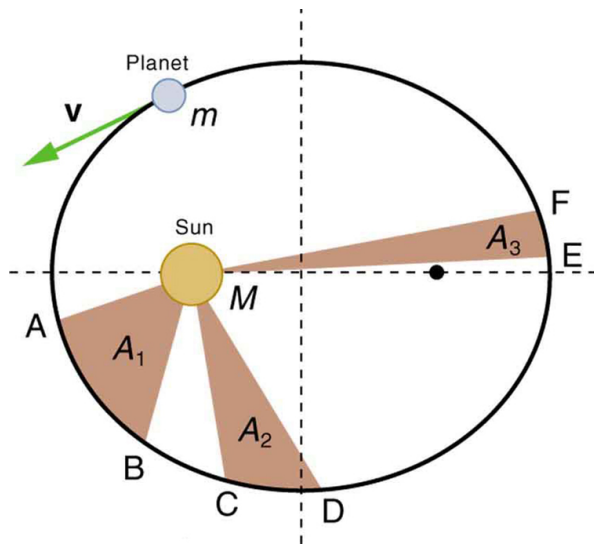
Kepler's Third Law

The ratio of the squares of the periods of any two planets about the Sun is equal to the ratio of the cubes of their average distances from the Sun. In equation form, this is

Equation:

$$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3},$$

where T is the period (time for one orbit) and r is the average radius. This equation is valid only for comparing two small masses orbiting the same large one. Most importantly, this is a descriptive equation only, giving no information as to the cause of the equality.



The shaded regions have equal areas. It takes equal times for m to go from A to B, from C to D, and from E to F. The mass m moves fastest when it is closest to M . Kepler's second law was originally devised for planets orbiting the Sun, but it has broader validity.

Note again that while, for historical reasons, Kepler's laws are stated for planets orbiting the Sun, they are actually valid for all bodies satisfying the two previously stated conditions.

Example:

Find the Time for One Orbit of an Earth Satellite

Given that the Moon orbits Earth each 27.3 d and that it is an average distance of 3.84×10^8 m from the center of Earth, calculate the period of an artificial satellite orbiting at an average altitude of 1500 km above Earth's surface.

Strategy

The period, or time for one orbit, is related to the radius of the orbit by Kepler's third law, given in mathematical form in $\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3}$. Let us use the subscript 1 for the Moon and the subscript 2 for the satellite. We are asked to find T_2 . The given information tells us that the orbital radius of the Moon is $r_1 = 3.84 \times 10^8$ m, and that the period of the Moon is $T_1 = 27.3$ d. The height of the artificial satellite above Earth's surface is given, and so we must add the radius of Earth (6380 km) to get $r_2 = (1500 + 6380)$ km = 7880 km. Now all quantities are known, and so T_2 can be found.

Solution

Kepler's third law is

Equation:

$$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3}.$$

To solve for T_2 , we cross-multiply and take the square root, yielding

Equation:

$$T_2^2 = T_1^2 \left(\frac{r_2}{r_1} \right)^3$$

Equation:

$$T_2 = T_1 \left(\frac{r_2}{r_1} \right)^{3/2}.$$

Substituting known values yields

Equation:

$$\begin{aligned} T_2 &= 27.3 \text{ d} \times \frac{24.0 \text{ h}}{\text{d}} \times \left(\frac{7880 \text{ km}}{3.84 \times 10^5 \text{ km}} \right)^{3/2} \\ &= 1.93 \text{ h.} \end{aligned}$$

Discussion This is a reasonable period for a satellite in a fairly low orbit. It is interesting that any satellite at this altitude will orbit in the same amount

of time. This fact is related to the condition that the satellite's mass is small compared with that of Earth.

People immediately search for deeper meaning when broadly applicable laws, like Kepler's, are discovered. It was Newton who took the next giant step when he proposed the law of universal gravitation. While Kepler was able to discover *what* was happening, Newton discovered that gravitational force was the cause.

Derivation of Kepler's Third Law for Circular Orbits

We shall derive Kepler's third law, starting with Newton's laws of motion and his universal law of gravitation. The point is to demonstrate that the force of gravity is the cause for Kepler's laws (although we will only derive the third one).

Let us consider a circular orbit of a small mass m around a large mass M , satisfying the two conditions stated at the beginning of this section. Gravity supplies the centripetal force to mass m . Starting with Newton's second law applied to circular motion,

Equation:

$$F_{\text{net}} = ma_c = m \frac{v^2}{r}.$$

The net external force on mass m is gravity, and so we substitute the force of gravity for F_{net} :

Equation:

$$G \frac{mM}{r^2} = m \frac{v^2}{r}.$$

The mass m cancels, yielding

Equation:

$$G \frac{M}{r} = v^2.$$

The fact that m cancels out is another aspect of the oft-noted fact that at a given location all masses fall with the same acceleration. Here we see that at a given orbital radius r , all masses orbit at the same speed. (This was implied by the result of the preceding worked example.) Now, to get at Kepler's third law, we must get the period T into the equation. By definition, period T is the time for one complete orbit. Now the average speed v is the circumference divided by the period—that is,

Equation:

$$v = \frac{2\pi r}{T}.$$

Substituting this into the previous equation gives

Equation:

$$G \frac{M}{r} = \frac{4\pi^2 r^2}{T^2}.$$

Solving for T^2 yields

Equation:

$$T^2 = \frac{4\pi^2}{GM} r^3.$$

Using subscripts 1 and 2 to denote two different satellites, and taking the ratio of the last equation for satellite 1 to satellite 2 yields

Equation:

$$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3}.$$

This is Kepler's third law. Note that Kepler's third law is valid only for comparing satellites of the same parent body, because only then does the mass of the parent body M cancel.

Now consider what we get if we solve $T^2 = \frac{4\pi^2}{GM} r^3$ for the ratio r^3/T^2 . We obtain a relationship that can be used to determine the mass M of a parent body from the orbits of its satellites:

Equation:

$$\frac{r^3}{T^2} = \frac{G}{4\pi^2} M.$$

If r and T are known for a satellite, then the mass M of the parent can be calculated. This principle has been used extensively to find the masses of heavenly bodies that have satellites. Furthermore, the ratio r^3/T^2 should be a constant for all satellites of the same parent body (because $r^3/T^2 = GM/4\pi^2$). (See [\[link\]](#)).

It is clear from [\[link\]](#) that the ratio of r^3/T^2 is constant, at least to the third digit, for all listed satellites of the Sun, and for those of Jupiter. Small variations in that ratio have two causes—uncertainties in the r and T data, and perturbations of the orbits due to other bodies. Interestingly, those perturbations can be—and have been—used to predict the location of new planets and moons. This is another verification of Newton's universal law of gravitation.

Note:

Making Connections

Newton's universal law of gravitation is modified by Einstein's general theory of relativity, as we shall see in [Particle Physics](#). Newton's gravity is not seriously in error—it was and still is an extremely good approximation for most situations. Einstein's modification is most noticeable in extremely large gravitational fields, such as near black holes. However, general relativity also explains such phenomena as small but long-known deviations of the orbit of the planet Mercury from classical predictions.

The Case for Simplicity

The development of the universal law of gravitation by Newton played a pivotal role in the history of ideas. While it is beyond the scope of this text to cover that history in any detail, we note some important points. The definition of planet set in 2006 by the International Astronomical Union (IAU) states that in the solar system, a planet is a celestial body that:

- 1. is in orbit around the Sun,
- 2. has sufficient mass to assume hydrostatic equilibrium and
- 3. has cleared the neighborhood around its orbit.

A non-satellite body fulfilling only the first two of the above criteria is classified as “dwarf planet.”

In 2006, Pluto was demoted to a ‘dwarf planet’ after scientists revised their definition of what constitutes a “true” planet.

Parent	Satellite	Average orbital radius $r(\text{km})$	Period $T(\text{y})$	$r^3 / T^2 (\text{km}^3 / \text{y}^2)$
Earth	Moon	3.84×10^5	0.07481	1.01×10^{19}
Sun	Mercury	5.79×10^7	0.2409	3.34×10^{24}

Parent	Satellite	Average orbital radius $r(\text{km})$	Period $T(\text{y})$	$r^3 / T^2 (\text{km}^3 / \text{y}^2)$
	Venus	1.082×10^8	0.6150	3.35×10^{24}
	Earth	1.496×10^8	1.000	3.35×10^{24}
	Mars	2.279×10^8	1.881	3.35×10^{24}
	Jupiter	7.783×10^8	11.86	3.35×10^{24}
	Saturn	1.427×10^9	29.46	3.35×10^{24}
	Neptune	4.497×10^9	164.8	3.35×10^{24}
	Pluto	5.90×10^9	248.3	3.33×10^{24}
Jupiter	Io	4.22×10^5	0.00485 (1.77 d)	3.19×10^{21}
	Europa	6.71×10^5	0.00972 (3.55 d)	3.20×10^{21}

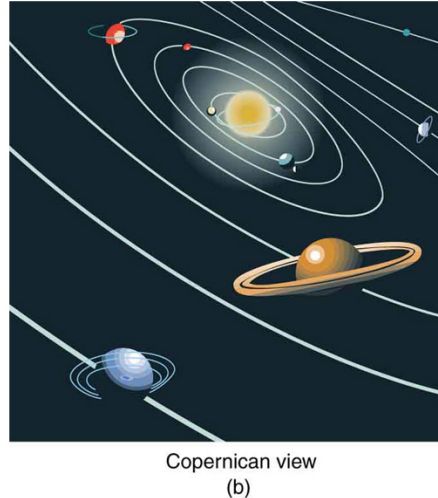
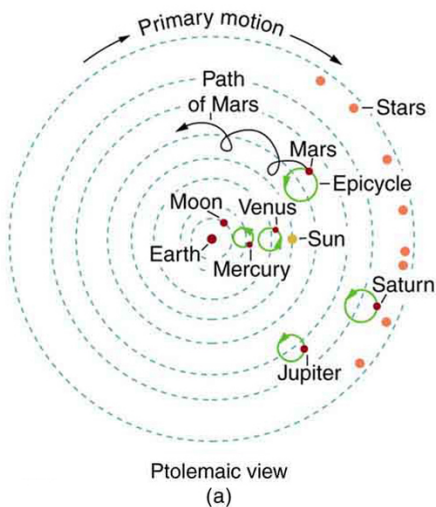
Parent	Satellite	Average orbital radius $r(\text{km})$	Period $T(\text{y})$	$r^3 / T^2 (\text{km}^3 / \text{y}^2)$
	Ganymede	1.07×10^6	0.0196 (7.16 d)	3.19×10^{21}
	Callisto	1.88×10^6	0.0457 (16.19 d)	3.20×10^{21}

Orbital Data and Kepler's Third Law

The universal law of gravitation is a good example of a physical principle that is very broadly applicable. That single equation for the gravitational force describes all situations in which gravity acts. It gives a cause for a vast number of effects, such as the orbits of the planets and moons in the solar system. It epitomizes the underlying unity and simplicity of physics.

Before the discoveries of Kepler, Copernicus, Galileo, Newton, and others, the solar system was thought to revolve around Earth as shown in [\[link\]](#)(a). This is called the Ptolemaic view, for the Greek philosopher who lived in the second century AD. This model is characterized by a list of facts for the motions of planets with no cause and effect explanation. There tended to be a different rule for each heavenly body and a general lack of simplicity.

[\[link\]](#)(b) represents the modern or Copernican model. In this model, a small set of rules and a single underlying force explain not only all motions in the solar system, but all other situations involving gravity. The breadth and simplicity of the laws of physics are compelling. As our knowledge of nature has grown, the basic simplicity of its laws has become ever more evident.



(a) The Ptolemaic model of the universe has Earth at the center with the Moon, the planets, the Sun, and the stars revolving about it in complex superpositions of circular paths. This geocentric model, which can be made progressively more accurate by adding more circles, is purely descriptive, containing no hints as to what are the causes of these motions. (b) The Copernican model has the Sun at the center of the solar system. It is fully explained by a small number of laws of physics, including Newton's universal law of gravitation.

Section Summary

- Kepler's laws are stated for a small mass m orbiting a larger mass M in near-isolation. Kepler's laws of planetary motion are then as follows:

Kepler's first law

The orbit of each planet about the Sun is an ellipse with the Sun at one focus.

Kepler's second law

Each planet moves so that an imaginary line drawn from the Sun to the planet sweeps out equal areas in equal times.

Kepler's third law

The ratio of the squares of the periods of any two planets about the Sun is equal to the ratio of the cubes of their average distances from the Sun:

Equation:

$$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3},$$

where T is the period (time for one orbit) and r is the average radius of the orbit.

- The period and radius of a satellite's orbit about a larger body M are related by

Equation:

$$T^2 = \frac{4\pi^2}{GM} r^3$$

or

Equation:

$$\frac{r^3}{T^2} = \frac{G}{4\pi^2} M.$$

Conceptual Questions

Exercise:

Problem:

In what frame(s) of reference are Kepler's laws valid? Are Kepler's laws purely descriptive, or do they contain causal information?

Problem Exercises

Exercise:

Problem:

A geosynchronous Earth satellite is one that has an orbital period of precisely 1 day. Such orbits are useful for communication and weather observation because the satellite remains above the same point on Earth (provided it orbits in the equatorial plane in the same direction as Earth's rotation). Calculate the radius of such an orbit based on the data for the moon in [\[link\]](#).

Exercise:

Problem:

Calculate the mass of the Sun based on data for Earth's orbit and compare the value obtained with the Sun's actual mass.

Solution:

$$1.98 \times 10^{30} \text{ kg}$$

Exercise:

Problem:

Find the mass of Jupiter based on data for the orbit of one of its moons, and compare your result with its actual mass.

Exercise:

Problem:

Find the ratio of the mass of Jupiter to that of Earth based on data in [\[link\]](#).

Solution:

$$\frac{M_J}{M_E} = 316$$

Exercise:

Problem:

Astronomical observations of our Milky Way galaxy indicate that it has a mass of about 8.0×10^{11} solar masses. A star orbiting on the galaxy's periphery is about 6.0×10^4 light years from its center. (a) What should the orbital period of that star be? (b) If its period is 6.0×10^7 years instead, what is the mass of the galaxy? Such calculations are used to imply the existence of "dark matter" in the universe and have indicated, for example, the existence of very massive black holes at the centers of some galaxies.

Exercise:**Problem: Integrated Concepts**

Space debris left from old satellites and their launchers is becoming a hazard to other satellites. (a) Calculate the speed of a satellite in an orbit 900 km above Earth's surface. (b) Suppose a loose rivet is in an orbit of the same radius that intersects the satellite's orbit at an angle of 90° relative to Earth. What is the velocity of the rivet relative to the satellite just before striking it? (c) Given the rivet is 3.00 mm in size, how long will its collision with the satellite last? (d) If its mass is 0.500 g, what is the average force it exerts on the satellite? (e) How much energy in joules is generated by the collision? (The satellite's velocity does not change appreciably, because its mass is much greater than the rivet's.)

Solution:

a) $7.4 \times 10^3 \text{ m/s}$

b) $1.05 \times 10^3 \text{ m/s}$

c) $2.86 \times 10^{-7} \text{ s}$

d) $1.84 \times 10^7 \text{ N}$

e) $2.76 \times 10^4 \text{ J}$

Exercise:

Problem: Unreasonable Results

(a) Based on Kepler's laws and information on the orbital characteristics of the Moon, calculate the orbital radius for an Earth satellite having a period of 1.00 h. (b) What is unreasonable about this result? (c) What is unreasonable or inconsistent about the premise of a 1.00 h orbit?

Solution:

a) $5.08 \times 10^3 \text{ km}$

b) This radius is unreasonable because it is less than the radius of earth.

c) The premise of a one-hour orbit is inconsistent with the known radius of the earth.

Exercise:**Problem: Construct Your Own Problem**

On February 14, 2000, the NEAR spacecraft was successfully inserted into orbit around Eros, becoming the first artificial satellite of an asteroid. Construct a problem in which you determine the orbital speed for a satellite near Eros. You will need to find the mass of the asteroid and consider such things as a safe distance for the orbit. Although Eros is not spherical, calculate the acceleration due to gravity on its surface at a point an average distance from its center of mass. Your instructor may also wish to have you calculate the escape velocity from this point on Eros.

Introduction to Work, Energy, and Energy Resources

class="introduction"

How many
forms of
energy can
you identify
in this
photograph
of a wind
farm in
Iowa?
(credit:
Jürgen from
Sandesneben
, Germany,
Wikimedia
Commons)



Energy plays an essential role both in everyday events and in scientific phenomena. You can no doubt name many forms of energy, from that provided by our foods, to the energy we use to run our cars, to the sunlight that warms us on the beach. You can also cite examples of what people call energy that may not be scientific, such as someone having an energetic personality. Not only does energy have many interesting forms, it is

involved in almost all phenomena, and is one of the most important concepts of physics. What makes it even more important is that the total amount of energy in the universe is constant. Energy can change forms, but it cannot appear from nothing or disappear without a trace. Energy is thus one of a handful of physical quantities that we say is *conserved*.

Conservation of energy (as physicists like to call the principle that energy can neither be created nor destroyed) is based on experiment. Even as scientists discovered new forms of energy, conservation of energy has always been found to apply. Perhaps the most dramatic example of this was supplied by Einstein when he suggested that mass is equivalent to energy (his famous equation $E = mc^2$).

From a societal viewpoint, energy is one of the major building blocks of modern civilization. Energy resources are key limiting factors to economic growth. The world use of energy resources, especially oil, continues to grow, with ominous consequences economically, socially, politically, and environmentally. We will briefly examine the world's energy use patterns at the end of this chapter.

There is no simple, yet accurate, scientific definition for energy. Energy is characterized by its many forms and the fact that it is conserved. We can loosely define **energy** as the ability to do work, admitting that in some circumstances not all energy is available to do work. Because of the association of energy with work, we begin the chapter with a discussion of work. Work is intimately related to energy and how energy moves from one system to another or changes form.

Work: The Scientific Definition

- Explain how an object must be displaced for a force on it to do work.
- Explain how relative directions of force and displacement determine whether the work done is positive, negative, or zero.

What It Means to Do Work

The scientific definition of work differs in some ways from its everyday meaning. Certain things we think of as hard work, such as writing an exam or carrying a heavy load on level ground, are not work as defined by a scientist. The scientific definition of work reveals its relationship to energy—whenever work is done, energy is transferred.

For work, in the scientific sense, to be done, a force must be exerted and there must be displacement in the direction of the force.

Formally, the **work** done on a system by a constant force is defined to be *the product of the component of the force in the direction of motion times the distance through which the force acts*. For one-way motion in one dimension, this is expressed in equation form as

Equation:

$$W = | \mathbf{F} | (\cos \theta) | \mathbf{d} |,$$

where W is work, \mathbf{d} is the displacement of the system, and θ is the angle between the force vector \mathbf{F} and the displacement vector \mathbf{d} , as in [\[link\]](#). We can also write this as

Equation:

$$W = Fd \cos \theta.$$

To find the work done on a system that undergoes motion that is not one-way or that is in two or three dimensions, we divide the motion into one-way one-dimensional segments and add up the work done over each segment.

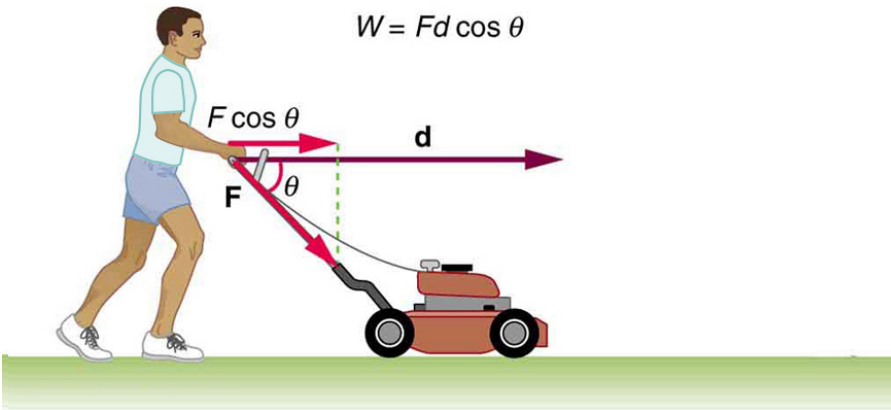
Note:**What is Work?**

The work done on a system by a constant force is *the product of the component of the force in the direction of motion times the distance through which the force acts*. For one-way motion in one dimension, this is expressed in equation form as

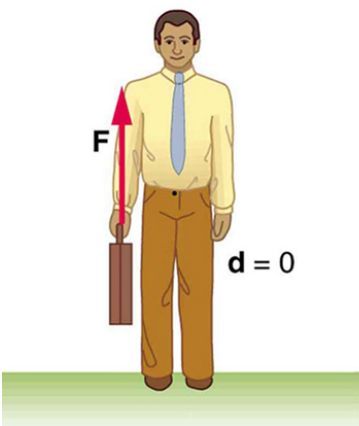
Equation:

$$W = Fd \cos \theta,$$

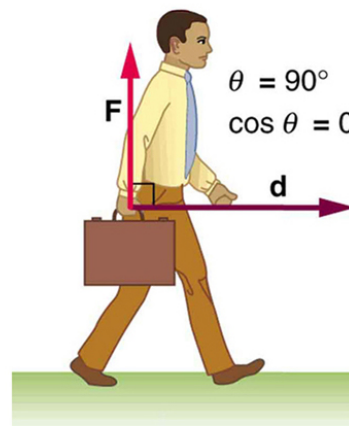
where W is work, F is the magnitude of the force on the system, d is the magnitude of the displacement of the system, and θ is the angle between the force vector \mathbf{F} and the displacement vector \mathbf{d} .



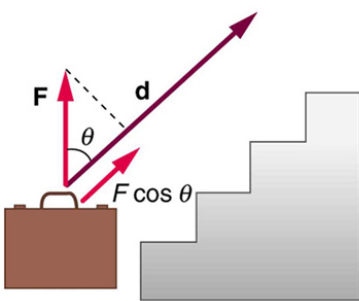
(a)



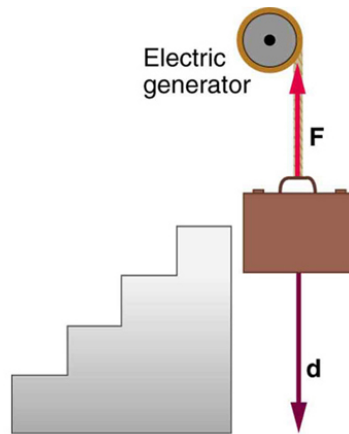
(b)



(c)



(d)



(e)

Examples of work. (a) The work done by the force \mathbf{F} on this lawn mower is $Fd \cos \theta$. Note that $F \cos \theta$ is the component of the force in the direction of motion. (b) A person holding a briefcase does no work on it, because there is no

displacement. No energy is transferred to or from the briefcase. (c) The person moving the briefcase horizontally at a constant speed does no work on it, and transfers no energy to it. (d) Work is done on the briefcase by carrying it up stairs at constant speed, because there is necessarily a component of force \mathbf{F} in the direction of the motion. Energy is transferred to the briefcase and could in turn be used to do work. (e) When the briefcase is lowered, energy is transferred out of the briefcase and into an electric generator. Here the work done on the briefcase by the generator is negative, removing energy from the briefcase, because \mathbf{F} and $d\mathbf{l}$ are in opposite directions.

To examine what the definition of work means, let us consider the other situations shown in [\[link\]](#). The person holding the briefcase in [\[link\]\(b\)](#) does no work, for example. Here $d = 0$, so $W = 0$. Why is it you get tired just holding a load? The answer is that your muscles are doing work against one another, *but they are doing no work on the system of interest* (the “briefcase-Earth system”—see [Gravitational Potential Energy](#) for more details). There must be displacement for work to be done, and there must be a component of the force in the direction of the motion. For example, the person carrying the briefcase on level ground in [\[link\]\(c\)](#) does no work on it, because the force is perpendicular to the motion. That is, $\cos 90^\circ = 0$, and so $W = 0$.

In contrast, when a force exerted on the system has a component in the direction of motion, such as in [\[link\]\(d\)](#), work is done—energy is transferred to the briefcase. Finally, in [\[link\]\(e\)](#), energy is transferred from the briefcase to a generator. There are two good ways to interpret this energy transfer. One interpretation is that the briefcase’s weight does work on the generator, giving it energy. The other interpretation is that the generator does negative work on the briefcase, thus removing energy from it. The drawing shows the latter, with the force from the generator upward

on the briefcase, and the displacement downward. This makes $\theta = 180^\circ$, and $\cos 180^\circ = -1$; therefore, W is negative.

Calculating Work

Work and energy have the same units. From the definition of work, we see that those units are force times distance. Thus, in SI units, work and energy are measured in **newton-meters**. A newton-meter is given the special name **joule** (J), and $1 \text{ J} = 1 \text{ N} \cdot \text{m} = 1 \text{ kg} \cdot \text{m}^2/\text{s}^2$. One joule is not a large amount of energy; it would lift a small 100-gram apple a distance of about 1 meter.

Example:

Calculating the Work You Do to Push a Lawn Mower Across a Large Lawn

How much work is done on the lawn mower by the person in [\[link\]](#)(a) if he exerts a constant force of 75.0 N at an angle 35° below the horizontal and pushes the mower 25.0 m on level ground? Convert the amount of work from joules to kilocalories and compare it with this person's average daily intake of 10,000 kJ (about 2400 kcal) of food energy. One *calorie* (1 cal) of heat is the amount required to warm 1 g of water by 1°C , and is equivalent to 4.184 J, while one *food calorie* (1 kcal) is equivalent to 4184 J.

Strategy

We can solve this problem by substituting the given values into the definition of work done on a system, stated in the equation $W = Fd \cos \theta$. The force, angle, and displacement are given, so that only the work W is unknown.

Solution

The equation for the work is

Equation:

$$W = Fd \cos \theta.$$

Substituting the known values gives

Equation:

$$\begin{aligned} W &= (75.0 \text{ N})(25.0 \text{ m}) \cos (35.0^\circ) \\ &= 1536 \text{ J} = 1.54 \times 10^3 \text{ J}. \end{aligned}$$

Converting the work in joules to kilocalories yields

$W = (1536 \text{ J})(1 \text{ kcal}/4184 \text{ J}) = 0.367 \text{ kcal}$. The ratio of the work done to the daily consumption is

Equation:

$$\frac{W}{2400 \text{ kcal}} = 1.53 \times 10^{-4}.$$

Discussion

This ratio is a tiny fraction of what the person consumes, but it is typical. Very little of the energy released in the consumption of food is used to do work. Even when we “work” all day long, less than 10% of our food energy intake is used to do work and more than 90% is converted to thermal energy or stored as chemical energy in fat.

Section Summary

- Work is the transfer of energy by a force acting on an object as it is displaced.
- The work W that a force \mathbf{F} does on an object is the product of the magnitude F of the force, times the magnitude d of the displacement, times the cosine of the angle θ between them. In symbols,

Equation:

$$W = Fd \cos \theta.$$

- The SI unit for work and energy is the joule (J), where $1 \text{ J} = 1 \text{ N} \cdot \text{m} = 1 \text{ kg} \cdot \text{m}^2/\text{s}^2$.
- The work done by a force is zero if the displacement is either zero or perpendicular to the force.

- The work done is positive if the force and displacement have the same direction, and negative if they have opposite direction.

Conceptual Questions

Exercise:

Problem:

Give an example of something we think of as work in everyday circumstances that is not work in the scientific sense. Is energy transferred or changed in form in your example? If so, explain how this is accomplished without doing work.

Exercise:

Problem:

Give an example of a situation in which there is a force and a displacement, but the force does no work. Explain why it does no work.

Exercise:

Problem:

Describe a situation in which a force is exerted for a long time but does no work. Explain.

Problems & Exercises

Exercise:

Problem:

How much work does a supermarket checkout attendant do on a can of soup he pushes 0.600 m horizontally with a force of 5.00 N? Express your answer in joules and kilocalories.

Solution:

Equation:

$$3.00 \text{ J} = 7.17 \times 10^{-4} \text{ kcal}$$

Exercise:**Problem:**

A 75.0-kg person climbs stairs, gaining 2.50 meters in height. Find the work done to accomplish this task.

Exercise:**Problem:**

(a) Calculate the work done on a 1500-kg elevator car by its cable to lift it 40.0 m at constant speed, assuming friction averages 100 N. (b) What is the work done on the lift by the gravitational force in this process? (c) What is the total work done on the lift?

Solution:

(a) $5.92 \times 10^5 \text{ J}$

(b) $-5.88 \times 10^5 \text{ J}$

(c) The net force is zero.

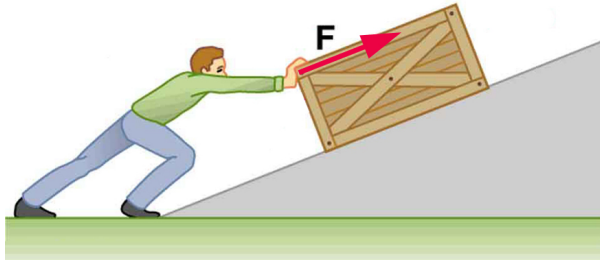
Exercise:**Problem:**

Suppose a car travels 108 km at a speed of 30.0 m/s, and uses 2.0 gal of gasoline. Only 30% of the gasoline goes into useful work by the force that keeps the car moving at constant speed despite friction. (See [\[link\]](#) for the energy content of gasoline.) (a) What is the magnitude of the force exerted to keep the car moving at constant speed? (b) If the required force is directly proportional to speed, how many gallons will be used to drive 108 km at a speed of 28.0 m/s?

Exercise:

Problem:

Calculate the work done by an 85.0-kg man who pushes a crate 4.00 m up along a ramp that makes an angle of 20.0° with the horizontal. (See [\[link\]](#).) He exerts a force of 500 N on the crate parallel to the ramp and moves at a constant speed. Be certain to include the work he does on the crate *and* on his body to get up the ramp.



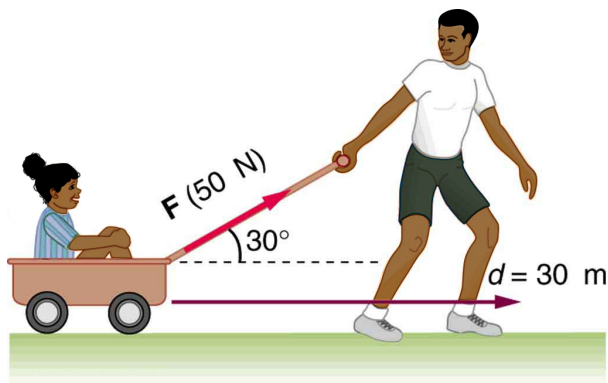
A man pushes a crate up a ramp.

Solution:**Equation:**

$$3.14 \times 10^3 \text{ J}$$

Exercise:**Problem:**

How much work is done by the boy pulling his sister 30.0 m in a wagon as shown in [\[link\]](#)? Assume no friction acts on the wagon.



The boy does work on the system of the wagon and the child when he pulls them as shown.

Exercise:

Problem:

A shopper pushes a grocery cart 20.0 m at constant speed on level ground, against a 35.0 N frictional force. He pushes in a direction 25.0° below the horizontal. (a) What is the work done on the cart by friction? (b) What is the work done on the cart by the gravitational force? (c) What is the work done on the cart by the shopper? (d) Find the force the shopper exerts, using energy considerations. (e) What is the total work done on the cart?

Solution:

(a) -700 J

(b) 0

(c) 700 J

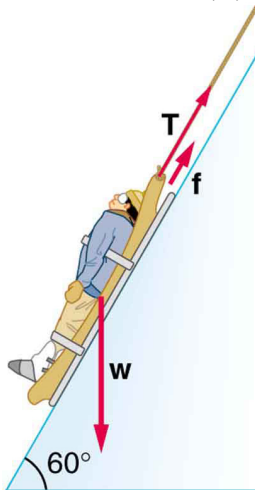
(d) 38.6 N

(e) 0

Exercise:

Problem:

Suppose the ski patrol lowers a rescue sled and victim, having a total mass of 90.0 kg , down a 60.0° slope at constant speed, as shown in [\[link\]](#). The coefficient of friction between the sled and the snow is 0.100 . (a) How much work is done by friction as the sled moves 30.0 m along the hill? (b) How much work is done by the rope on the sled in this distance? (c) What is the work done by the gravitational force on the sled? (d) What is the total work done?



A rescue sled and victim are lowered down a steep slope.

Glossary

energy

the ability to do work

work

the transfer of energy by a force that causes an object to be displaced;
the product of the component of the force in the direction of the
displacement and the magnitude of the displacement

joule

SI unit of work and energy, equal to one newton-meter

Kinetic Energy and the Work-Energy Theorem

- Explain work as a transfer of energy and net work as the work done by the net force.
- Explain and apply the work-energy theorem.

Work Transfers Energy

What happens to the work done on a system? Energy is transferred into the system, but in what form? Does it remain in the system or move on? The answers depend on the situation. For example, if the lawn mower in [\[link\]](#) (a) is pushed just hard enough to keep it going at a constant speed, then energy put into the mower by the person is removed continuously by friction, and eventually leaves the system in the form of heat transfer. In contrast, work done on the briefcase by the person carrying it up stairs in [\[link\]](#) (d) is stored in the briefcase-Earth system and can be recovered at any time, as shown in [\[link\]](#) (e). In fact, the building of the pyramids in ancient Egypt is an example of storing energy in a system by doing work on the system. Some of the energy imparted to the stone blocks in lifting them during construction of the pyramids remains in the stone-Earth system and has the potential to do work.

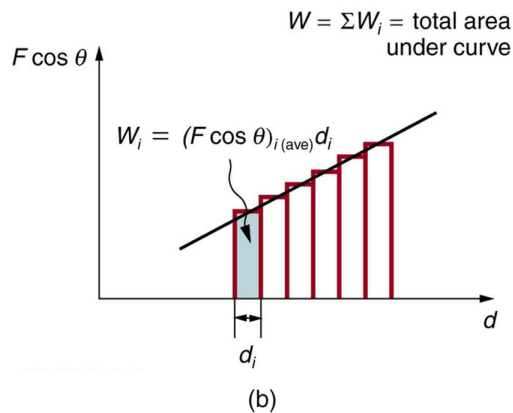
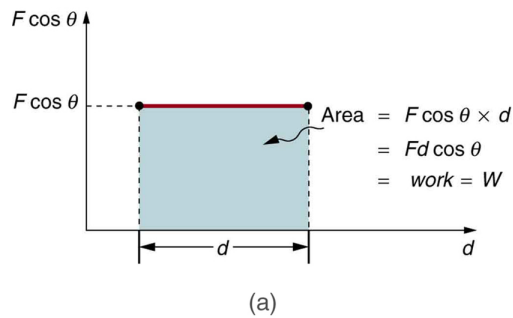
In this section we begin the study of various types of work and forms of energy. We will find that some types of work leave the energy of a system constant, for example, whereas others change the system in some way, such as making it move. We will also develop definitions of important forms of energy, such as the energy of motion.

Net Work and the Work-Energy Theorem

We know from the study of Newton's laws in [Dynamics: Force and Newton's Laws of Motion](#) that net force causes acceleration. We will see in this section that work done by the net force gives a system energy of motion, and in the process we will also find an expression for the energy of motion.

Let us start by considering the total, or net, work done on a system. Net work is defined to be the sum of work done by all external forces—that is, **net work** is the work done by the net external force \mathbf{F}_{net} . In equation form, this is $W_{\text{net}} = F_{\text{net}}d \cos \theta$ where θ is the angle between the force vector and the displacement vector.

[\[link\]](#)(a) shows a graph of force versus displacement for the component of the force in the direction of the displacement—that is, an $F \cos \theta$ vs. d graph. In this case, $F \cos \theta$ is constant. You can see that the area under the graph is $Fd \cos \theta$, or the work done. [\[link\]](#)(b) shows a more general process where the force varies. The area under the curve is divided into strips, each having an average force $(F \cos \theta)_{i(\text{ave})}$. The work done is $(F \cos \theta)_{i(\text{ave})}d_i$ for each strip, and the total work done is the sum of the W_i . Thus the total work done is the total area under the curve, a useful property to which we shall refer later.

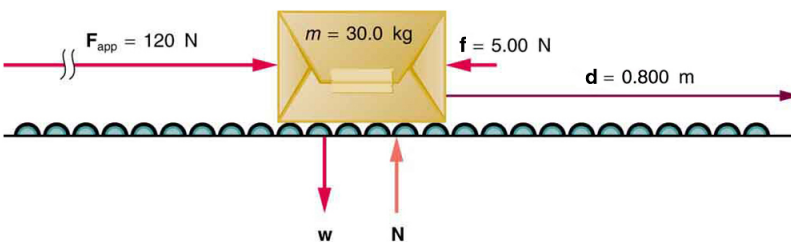


(a) A graph of $F \cos \theta$ vs. d , when $F \cos \theta$ is

constant. The area under the curve represents the work done by the force.

(b) A graph of $F \cos \theta$ vs. d in which the force varies. The work done for each interval is the area of each strip; thus, the total area under the curve equals the total work done.

Net work will be simpler to examine if we consider a one-dimensional situation where a force is used to accelerate an object in a direction parallel to its initial velocity. Such a situation occurs for the package on the roller belt conveyor system shown in [\[link\]](#).



A package on a roller belt is pushed horizontally through a distance d .

The force of gravity and the normal force acting on the package are perpendicular to the displacement and do no work. Moreover, they are also equal in magnitude and opposite in direction so they cancel in calculating the net force. The net force arises solely from the horizontal applied force F_{app} and the horizontal friction force f . Thus, as expected, the net force is

parallel to the displacement, so that $\theta = 0^\circ$ and $\cos \theta = 1$, and the net work is given by

Equation:

$$W_{\text{net}} = F_{\text{net}}d.$$

The effect of the net force \mathbf{F}_{net} is to accelerate the package from v_0 to v . The kinetic energy of the package increases, indicating that the net work done on the system is positive. (See [\[link\]](#).) By using Newton's second law, and doing some algebra, we can reach an interesting conclusion. Substituting $F_{\text{net}} = ma$ from Newton's second law gives

Equation:

$$W_{\text{net}} = mad.$$

To get a relationship between net work and the speed given to a system by the net force acting on it, we take $d = x - x_0$ and use the equation studied in [Motion Equations for Constant Acceleration in One Dimension](#) for the change in speed over a distance d if the acceleration has the constant value a ; namely, $v^2 = v_0^2 + 2ad$ (note that a appears in the expression for the net work). Solving for acceleration gives $a = \frac{v^2 - v_0^2}{2d}$. When a is substituted into the preceding expression for W_{net} , we obtain

Equation:

$$W_{\text{net}} = m \left(\frac{v^2 - v_0^2}{2d} \right) d.$$

The d cancels, and we rearrange this to obtain

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2.$$

This expression is called the **work-energy theorem**, and it actually applies *in general* (even for forces that vary in direction and magnitude), although we have derived it for the special case of a constant force parallel to the displacement. The theorem implies that the net work on a system equals the change in the quantity $\frac{1}{2}mv^2$. This quantity is our first example of a form of energy.

Note:

The Work-Energy Theorem

The net work on a system equals the change in the quantity $\frac{1}{2}mv^2$.

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$$

The quantity $\frac{1}{2}mv^2$ in the work-energy theorem is defined to be the translational **kinetic energy** (KE) of a mass m moving at a speed v . (*Translational* kinetic energy is distinct from *rotational* kinetic energy, which is considered later.) In equation form, the translational kinetic energy,

Equation:

$$\text{KE} = \frac{1}{2}mv^2,$$

is the energy associated with translational motion. Kinetic energy is a form of energy associated with the motion of a particle, single body, or system of objects moving together.

We are aware that it takes energy to get an object, like a car or the package in [\[link\]](#), up to speed, but it may be a bit surprising that kinetic energy is proportional to speed squared. This proportionality means, for example, that a car traveling at 100 km/h has four times the kinetic energy it has at 50

km/h, helping to explain why high-speed collisions are so devastating. We will now consider a series of examples to illustrate various aspects of work and energy.

Example:

Calculating the Kinetic Energy of a Package

Suppose a 30.0-kg package on the roller belt conveyor system in [\[link\]](#) is moving at 0.500 m/s. What is its kinetic energy?

Strategy

Because the mass m and speed v are given, the kinetic energy can be calculated from its definition as given in the equation $\text{KE} = \frac{1}{2}mv^2$.

Solution

The kinetic energy is given by

Equation:

$$\text{KE} = \frac{1}{2}mv^2.$$

Entering known values gives

Equation:

$$\text{KE} = 0.5(30.0 \text{ kg})(0.500 \text{ m/s})^2,$$

which yields

Equation:

$$\text{KE} = 3.75 \text{ kg} \cdot \text{m}^2/\text{s}^2 = 3.75 \text{ J}.$$

Discussion

Note that the unit of kinetic energy is the joule, the same as the unit of work, as mentioned when work was first defined. It is also interesting that, although this is a fairly massive package, its kinetic energy is not large at this relatively low speed. This fact is consistent with the observation that people can move packages like this without exhausting themselves.

Example:**Determining the Work to Accelerate a Package**

Suppose that you push on the 30.0-kg package in [\[link\]](#) with a constant force of 120 N through a distance of 0.800 m, and that the opposing friction force averages 5.00 N.

(a) Calculate the net work done on the package. (b) Solve the same problem as in part (a), this time by finding the work done by each force that contributes to the net force.

Strategy and Concept for (a)

This is a motion in one dimension problem, because the downward force (from the weight of the package) and the normal force have equal magnitude and opposite direction, so that they cancel in calculating the net force, while the applied force, friction, and the displacement are all horizontal. (See [\[link\]](#).) As expected, the net work is the net force times distance.

Solution for (a)

The net force is the push force minus friction, or

$F_{\text{net}} = 120 \text{ N} - 5.00 \text{ N} = 115 \text{ N}$. Thus the net work is

Equation:

$$\begin{aligned} W_{\text{net}} &= F_{\text{net}}d = (115 \text{ N})(0.800 \text{ m}) \\ &= 92.0 \text{ N} \cdot \text{m} = 92.0 \text{ J.} \end{aligned}$$

Discussion for (a)

This value is the net work done on the package. The person actually does more work than this, because friction opposes the motion. Friction does negative work and removes some of the energy the person expends and converts it to thermal energy. The net work equals the sum of the work done by each individual force.

Strategy and Concept for (b)

The forces acting on the package are gravity, the normal force, the force of friction, and the applied force. The normal force and force of gravity are each perpendicular to the displacement, and therefore do no work.

Solution for (b)

The applied force does work.

Equation:

$$\begin{aligned}
 W_{\text{app}} &= F_{\text{app}} d \cos(0^\circ) = F_{\text{app}} d \\
 &= (120 \text{ N})(0.800 \text{ m}) \\
 &= 96.0 \text{ J}
 \end{aligned}$$

The friction force and displacement are in opposite directions, so that $\theta = 180^\circ$, and the work done by friction is

Equation:

$$\begin{aligned}
 W_{\text{fr}} &= F_{\text{fr}} d \cos(180^\circ) = -F_{\text{fr}} d \\
 &= -(5.00 \text{ N})(0.800 \text{ m}) \\
 &= -4.00 \text{ J.}
 \end{aligned}$$

So the amounts of work done by gravity, by the normal force, by the applied force, and by friction are, respectively,

Equation:

$$\begin{aligned}
 W_{\text{gr}} &= 0, \\
 W_{\text{N}} &= 0, \\
 W_{\text{app}} &= 96.0 \text{ J}, \\
 W_{\text{fr}} &= -4.00 \text{ J.}
 \end{aligned}$$

The total work done as the sum of the work done by each force is then seen to be

Equation:

$$W_{\text{total}} = W_{\text{gr}} + W_{\text{N}} + W_{\text{app}} + W_{\text{fr}} = 92.0 \text{ J.}$$

Discussion for (b)

The calculated total work W_{total} as the sum of the work by each force agrees, as expected, with the work W_{net} done by the net force. The work done by a collection of forces acting on an object can be calculated by either approach.

Example:

Determining Speed from Work and Energy

Find the speed of the package in [\[link\]](#) at the end of the push, using work and energy concepts.

Strategy

Here the work-energy theorem can be used, because we have just calculated the net work, W_{net} , and the initial kinetic energy, $\frac{1}{2}mv_0^2$. These calculations allow us to find the final kinetic energy, $\frac{1}{2}mv^2$, and thus the final speed v .

Solution

The work-energy theorem in equation form is

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2.$$

Solving for $\frac{1}{2}mv^2$ gives

Equation:

$$\frac{1}{2}mv^2 = W_{\text{net}} + \frac{1}{2}mv_0^2.$$

Thus,

Equation:

$$\frac{1}{2}mv^2 = 92.0 \text{ J} + 3.75 \text{ J} = 95.75 \text{ J}.$$

Solving for the final speed as requested and entering known values gives

Equation:

$$\begin{aligned} v &= \sqrt{\frac{2(95.75 \text{ J})}{m}} = \sqrt{\frac{191.5 \text{ kg}\cdot\text{m}^2/\text{s}^2}{30.0 \text{ kg}}} \\ &= 2.53 \text{ m/s}. \end{aligned}$$

Discussion

Using work and energy, we not only arrive at an answer, we see that the final kinetic energy is the sum of the initial kinetic energy and the net work

done on the package. This means that the work indeed adds to the energy of the package.

Example:

Work and Energy Can Reveal Distance, Too

How far does the package in [\[link\]](#) coast after the push, assuming friction remains constant? Use work and energy considerations.

Strategy

We know that once the person stops pushing, friction will bring the package to rest. In terms of energy, friction does negative work until it has removed all of the package's kinetic energy. The work done by friction is the force of friction times the distance traveled times the cosine of the angle between the friction force and displacement; hence, this gives us a way of finding the distance traveled after the person stops pushing.

Solution

The normal force and force of gravity cancel in calculating the net force. The horizontal friction force is then the net force, and it acts opposite to the displacement, so $\theta = 180^\circ$. To reduce the kinetic energy of the package to zero, the work W_{fr} by friction must be minus the kinetic energy that the package started with plus what the package accumulated due to the pushing. Thus $W_{\text{fr}} = -95.75 \text{ J}$. Furthermore, $W_{\text{fr}} = f d' \cos \theta = -f d'$, where d' is the distance it takes to stop. Thus,

Equation:

$$d' = -\frac{W_{\text{fr}}}{f} = -\frac{-95.75 \text{ J}}{5.00 \text{ N}},$$

and so

Equation:

$$d' = 19.2 \text{ m}.$$

Discussion

This is a reasonable distance for a package to coast on a relatively friction-free conveyor system. Note that the work done by friction is negative (the

force is in the opposite direction of motion), so it removes the kinetic energy.

Some of the examples in this section can be solved without considering energy, but at the expense of missing out on gaining insights about what work and energy are doing in this situation. On the whole, solutions involving energy are generally shorter and easier than those using kinematics and dynamics alone.

Section Summary

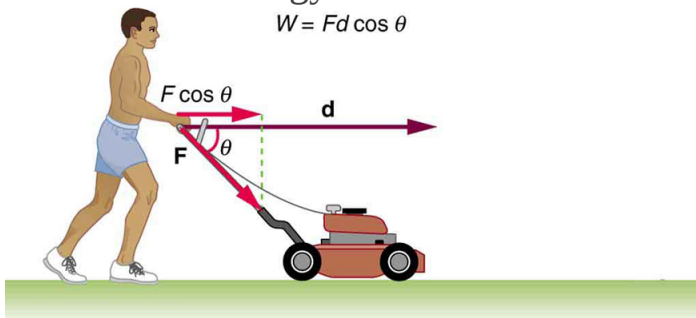
- The net work W_{net} is the work done by the net force acting on an object.
- Work done on an object transfers energy to the object.
- The translational kinetic energy of an object of mass m moving at speed v is $\text{KE} = \frac{1}{2}mv^2$.
- The work-energy theorem states that the net work W_{net} on a system changes its kinetic energy, $W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$.

Conceptual Questions

Exercise:

Problem:

The person in [\[link\]](#) does work on the lawn mower. Under what conditions would the mower gain energy? Under what conditions would it lose energy?



Exercise:**Problem:**

Work done on a system puts energy into it. Work done by a system removes energy from it. Give an example for each statement.

Exercise:**Problem:**

When solving for speed in [\[link\]](#), we kept only the positive root. Why?

Problems & Exercises**Exercise:****Problem:**

Compare the kinetic energy of a 20,000-kg truck moving at 110 km/h with that of an 80.0-kg astronaut in orbit moving at 27,500 km/h.

Solution:

1/250

Exercise:**Problem:**

(a) How fast must a 3000-kg elephant move to have the same kinetic energy as a 65.0-kg sprinter running at 10.0 m/s? (b) Discuss how the larger energies needed for the movement of larger animals would relate to metabolic rates.

Exercise:**Problem:**

Confirm the value given for the kinetic energy of an aircraft carrier in [\[link\]](#). You will need to look up the definition of a nautical mile (1 knot = 1 nautical mile/h).

Solution:

$$1.1 \times 10^{10} \text{ J}$$

Exercise:**Problem:**

(a) Calculate the force needed to bring a 950-kg car to rest from a speed of 90.0 km/h in a distance of 120 m (a fairly typical distance for a non-panic stop). (b) Suppose instead the car hits a concrete abutment at full speed and is brought to a stop in 2.00 m. Calculate the force exerted on the car and compare it with the force found in part (a).

Exercise:**Problem:**

A car's bumper is designed to withstand a 4.0-km/h (1.1-m/s) collision with an immovable object without damage to the body of the car. The bumper cushions the shock by absorbing the force over a distance. Calculate the magnitude of the average force on a bumper that collapses 0.200 m while bringing a 900-kg car to rest from an initial speed of 1.1 m/s.

Solution:

$$2.8 \times 10^3 \text{ N}$$

Exercise:

Problem:

Boxing gloves are padded to lessen the force of a blow. (a) Calculate the force exerted by a boxing glove on an opponent's face, if the glove and face compress 7.50 cm during a blow in which the 7.00-kg arm and glove are brought to rest from an initial speed of 10.0 m/s. (b) Calculate the force exerted by an identical blow in the gory old days when no gloves were used and the knuckles and face would compress only 2.00 cm. (c) Discuss the magnitude of the force with glove on. Does it seem high enough to cause damage even though it is lower than the force with no glove?

Exercise:**Problem:**

Using energy considerations, calculate the average force a 60.0-kg sprinter exerts backward on the track to accelerate from 2.00 to 8.00 m/s in a distance of 25.0 m, if he encounters a headwind that exerts an average force of 30.0 N against him.

Solution:

102 N

Glossary

net work

work done by the net force, or vector sum of all the forces, acting on an object

work-energy theorem

the result, based on Newton's laws, that the net work done on an object is equal to its change in kinetic energy

kinetic energy

the energy an object has by reason of its motion, equal to $\frac{1}{2}mv^2$ for the translational (i.e., non-rotational) motion of an object of mass m moving at speed v

Gravitational Potential Energy

- Explain gravitational potential energy in terms of work done against gravity.
- Show that the gravitational potential energy of an object of mass m at height h on Earth is given by $PE_g = mgh$.
- Show how knowledge of the potential energy as a function of position can be used to simplify calculations and explain physical phenomena.

Work Done Against Gravity

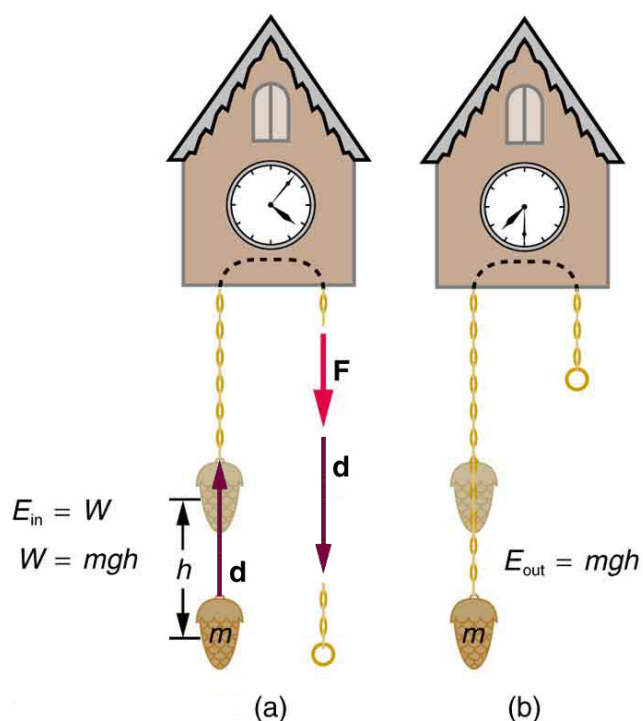
Climbing stairs and lifting objects is work in both the scientific and everyday sense—it is work done against the gravitational force. When there is work, there is a transformation of energy. The work done against the gravitational force goes into an important form of stored energy that we will explore in this section.

Let us calculate the work done in lifting an object of mass m through a height h , such as in [\[link\]](#). If the object is lifted straight up at constant speed, then the force needed to lift it is equal to its weight mg . The work done on the mass is then $W = Fd = mgh$. We define this to be the **gravitational potential energy** (PE_g) put into (or gained by) the object-Earth system. This energy is associated with the state of separation between two objects that attract each other by the gravitational force. For convenience, we refer to this as the PE_g gained by the object, recognizing that this is energy stored in the gravitational field of Earth. Why do we use the word “system”? Potential energy is a property of a system rather than of a single object—due to its physical position. An object’s gravitational potential is due to its position relative to the surroundings within the Earth-object system. The force applied to the object is an external force, from outside the system. When it does positive work it increases the gravitational potential energy of the system. Because gravitational potential energy depends on relative position, we need a reference level at which to set the potential energy equal to 0. We usually choose this point to be Earth’s surface, but this point is arbitrary; what is important is the *difference* in gravitational potential energy, because this difference is what relates to the work done. The difference in gravitational potential energy of an object (in the Earth-object system) between two rungs of a ladder will be the same for the first two rungs as for the last two rungs.

Converting Between Potential Energy and Kinetic Energy

Gravitational potential energy may be converted to other forms of energy, such as kinetic energy. If we release the mass, gravitational force will do an amount of work

equal to mgh on it, thereby increasing its kinetic energy by that same amount (by the work-energy theorem). We will find it more useful to consider just the conversion of PE_g to KE without explicitly considering the intermediate step of work. (See [\[link\]](#).) This shortcut makes it is easier to solve problems using energy (if possible) rather than explicitly using forces.



(a) The work done to lift the weight is stored in the mass-Earth system as gravitational potential energy. (b) As the weight moves downward, this gravitational potential energy is transferred to the cuckoo clock.

More precisely, we define the *change* in gravitational potential energy ΔPE_g to be
Equation:

$$\Delta PE_g = mgh,$$

where, for simplicity, we denote the change in height by h rather than the usual Δh . Note that h is positive when the final height is greater than the initial height, and vice versa. For example, if a 0.500-kg mass hung from a cuckoo clock is raised 1.00 m, then its change in gravitational potential energy is

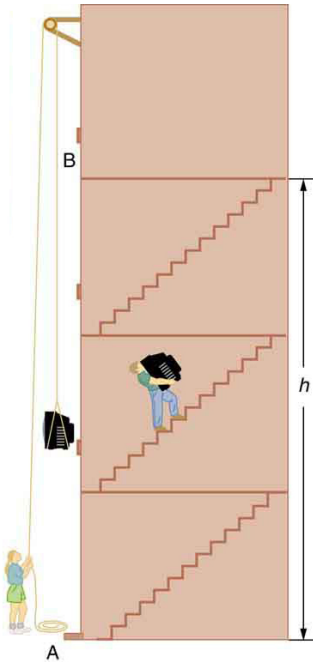
Equation:

$$\begin{aligned} mgh &= (0.500 \text{ kg}) (9.80 \text{ m/s}^2) (1.00 \text{ m}) \\ &= 4.90 \text{ kg} \cdot \text{m}^2/\text{s}^2 = 4.90 \text{ J}. \end{aligned}$$

Note that the units of gravitational potential energy turn out to be joules, the same as for work and other forms of energy. As the clock runs, the mass is lowered. We can think of the mass as gradually giving up its 4.90 J of gravitational potential energy, *without directly considering the force of gravity that does the work*.

Using Potential Energy to Simplify Calculations

The equation $\Delta \text{PE}_g = mgh$ applies for any path that has a change in height of h , not just when the mass is lifted straight up. (See [\[link\]](#).) It is much easier to calculate mgh (a simple multiplication) than it is to calculate the work done along a complicated path. The idea of gravitational potential energy has the double advantage that it is very broadly applicable and it makes calculations easier. From now on, we will consider that any change in vertical position h of a mass m is accompanied by a change in gravitational potential energy mgh , and we will avoid the equivalent but more difficult task of calculating work done by or against the gravitational force.



The change in
gravitational
potential energy
(ΔPE_g)
between points
A and B is
independent of
the path.

$\Delta PE_g = mgh$
for any path
between the two
points. Gravity
is one of a small
class of forces
where the work
done by or
against the force
depends only on
the starting and
ending points,
not on the path
between them.

Example:**The Force to Stop Falling**

A 60.0-kg person jumps onto the floor from a height of 3.00 m. If he lands stiffly (with his knee joints compressing by 0.500 cm), calculate the force on the knee joints.

Strategy

This person's energy is brought to zero in this situation by the work done on him by the floor as he stops. The initial PE_g is transformed into KE as he falls. The work done by the floor reduces this kinetic energy to zero.

Solution

The work done on the person by the floor as he stops is given by

Equation:

$$W = Fd \cos \theta = -Fd,$$

with a minus sign because the displacement while stopping and the force from floor are in opposite directions ($\cos \theta = \cos 180^\circ = -1$). The floor removes energy from the system, so it does negative work.

The kinetic energy the person has upon reaching the floor is the amount of potential energy lost by falling through height h :

Equation:

$$KE = -\Delta PE_g = -mgh,$$

The distance d that the person's knees bend is much smaller than the height h of the fall, so the additional change in gravitational potential energy during the knee bend is ignored.

The work W done by the floor on the person stops the person and brings the person's kinetic energy to zero:

Equation:

$$W = -KE = mgh.$$

Combining this equation with the expression for W gives

Equation:

$$-Fd = mgh.$$

Recalling that h is negative because the person fell *down*, the force on the knee joints is given by

Equation:

$$F = -\frac{mgh}{d} = -\frac{(60.0 \text{ kg})(9.80 \text{ m/s}^2)(-3.00 \text{ m})}{5.00 \times 10^{-3} \text{ m}} = 3.53 \times 10^5 \text{ N}.$$

Discussion

Such a large force (500 times more than the person's weight) over the short impact time is enough to break bones. A much better way to cushion the shock is by bending the legs or rolling on the ground, increasing the time over which the force acts. A bending motion of 0.5 m this way yields a force 100 times smaller than in the example. A kangaroo's hopping shows this method in action. The kangaroo is the only large animal to use hopping for locomotion, but the shock in hopping is cushioned by the bending of its hind legs in each jump.(See [\[link\]](#).)

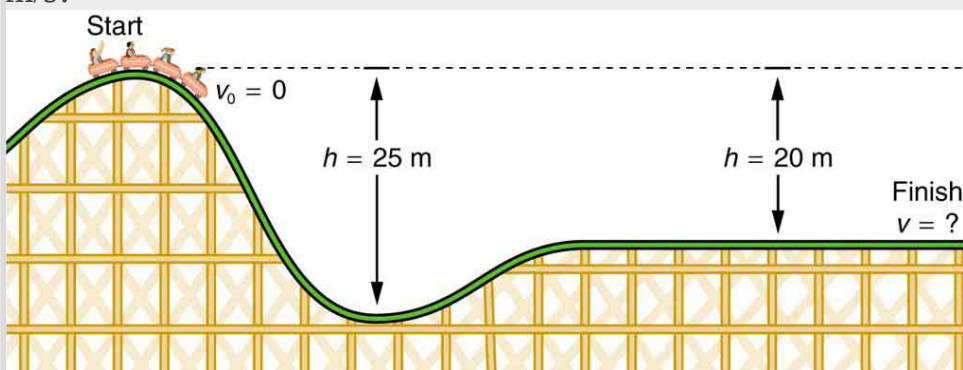


The work done by the ground upon the kangaroo reduces its kinetic energy to zero as it lands. However, by applying the force of the ground on the hind legs over a longer distance, the impact on the bones is reduced.
(credit: Chris Samuel, Flickr)

Example:

Finding the Speed of a Roller Coaster from its Height

(a) What is the final speed of the roller coaster shown in [\[link\]](#) if it starts from rest at the top of the 20.0 m hill and work done by frictional forces is negligible? (b) What is its final speed (again assuming negligible friction) if its initial speed is 5.00 m/s?



The speed of a roller coaster increases as gravity pulls it downhill and is greatest at its lowest point. Viewed in terms of energy, the roller-coaster-Earth system's gravitational potential energy is converted to kinetic energy. If work done by friction is negligible, all ΔPE_g is converted to KE.

Strategy

The roller coaster loses potential energy as it goes downhill. We neglect friction, so that the remaining force exerted by the track is the normal force, which is perpendicular to the direction of motion and does no work. The net work on the roller coaster is then done by gravity alone. The *loss* of gravitational potential energy from moving *downward* through a distance h equals the *gain* in kinetic energy. This can be written in equation form as $-\Delta PE_g = \Delta KE$. Using the equations for PE_g and KE, we can solve for the final speed v , which is the desired quantity.

Solution for (a)

Here the initial kinetic energy is zero, so that $\Delta KE = \frac{1}{2}mv^2$. The equation for change in potential energy states that $\Delta PE_g = mgh$. Since h is negative in this case, we will rewrite this as $\Delta PE_g = -mg |h|$ to show the minus sign clearly. Thus,

Equation:

$$-\Delta PE_g = \Delta KE$$

becomes

Equation:

$$mg | h | = \frac{1}{2}mv^2.$$

Solving for v , we find that mass cancels and that

Equation:

$$v = \sqrt{2g | h |}.$$

Substituting known values,

Equation:

$$\begin{aligned} v &= \sqrt{2(9.80 \text{ m/s}^2)(20.0 \text{ m})} \\ &= 19.8 \text{ m/s.} \end{aligned}$$

Solution for (b)

Again $-\Delta PE_g = \Delta KE$. In this case there is initial kinetic energy, so

$\Delta KE = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$. Thus,

Equation:

$$mg | h | = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2.$$

Rearranging gives

Equation:

$$\frac{1}{2}mv^2 = mg | h | + \frac{1}{2}mv_0^2.$$

This means that the final kinetic energy is the sum of the initial kinetic energy and the gravitational potential energy. Mass again cancels, and

Equation:

$$v = \sqrt{2g | h | + v_0^2}.$$

This equation is very similar to the kinematics equation $v = \sqrt{v_0^2 + 2ad}$, but it is more general—the kinematics equation is valid only for constant acceleration, whereas our equation above is valid for any path regardless of whether the object moves with a constant acceleration. Now, substituting known values gives

Equation:

$$\begin{aligned} v &= \sqrt{2(9.80 \text{ m/s}^2)(20.0 \text{ m}) + (5.00 \text{ m/s})^2} \\ &= 20.4 \text{ m/s.} \end{aligned}$$

Discussion and Implications

First, note that mass cancels. This is quite consistent with observations made in [Falling Objects](#) that all objects fall at the same rate if friction is negligible. Second, only the speed of the roller coaster is considered; there is no information about its direction at any point. This reveals another general truth. When friction is negligible, the speed of a falling body depends only on its initial speed and height, and not on its mass or the path taken. For example, the roller coaster will have the same final speed whether it falls 20.0 m straight down or takes a more complicated path like the one in the figure. Third, and perhaps unexpectedly, the final speed in part (b) is greater than in part (a), but by far less than 5.00 m/s. Finally, note that speed can be found at *any* height along the way by simply using the appropriate value of h at the point of interest.

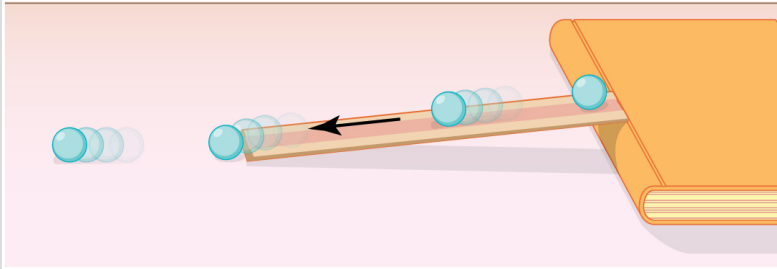
We have seen that work done by or against the gravitational force depends only on the starting and ending points, and not on the path between, allowing us to define the simplifying concept of gravitational potential energy. We can do the same thing for a few other forces, and we will see that this leads to a formal definition of the law of conservation of energy.

Note:

Making Connections: Take-Home Investigation—Converting Potential to Kinetic Energy

One can study the conversion of gravitational potential energy into kinetic energy in this experiment. On a smooth, level surface, use a ruler of the kind that has a groove running along its length and a book to make an incline (see [link](#)). Place a marble at the 10-cm position on the ruler and let it roll down the ruler. When it hits the level surface, measure the time it takes to roll one meter. Now place the marble

at the 20-cm and the 30-cm positions and again measure the times it takes to roll 1 m on the level surface. Find the velocity of the marble on the level surface for all three positions. Plot velocity squared versus the distance traveled by the marble. What is the shape of each plot? If the shape is a straight line, the plot shows that the marble's kinetic energy at the bottom is proportional to its potential energy at the release point.



A marble rolls down a ruler, and its speed on the level surface is measured.

Section Summary

- Work done against gravity in lifting an object becomes potential energy of the object-Earth system.
- The change in gravitational potential energy, ΔPE_g , is $\Delta PE_g = mgh$, with h being the increase in height and g the acceleration due to gravity.
- The gravitational potential energy of an object near Earth's surface is due to its position in the mass-Earth system. Only differences in gravitational potential energy, ΔPE_g , have physical significance.
- As an object descends without friction, its gravitational potential energy changes into kinetic energy corresponding to increasing speed, so that $\Delta KE = -\Delta PE_g$.

Conceptual Questions

Exercise:

Problem:

In [\[link\]](#), we calculated the final speed of a roller coaster that descended 20 m in height and had an initial speed of 5 m/s downhill. Suppose the roller coaster had had an initial speed of 5 m/s *uphill* instead, and it coasted uphill, stopped, and then rolled back down to a final point 20 m below the start. We would find in that case that its final speed is the same as its initial speed. Explain in terms of conservation of energy.

Exercise:**Problem:**

Does the work you do on a book when you lift it onto a shelf depend on the path taken? On the time taken? On the height of the shelf? On the mass of the book?

Problems & Exercises**Exercise:****Problem:**

A hydroelectric power facility (see [\[link\]](#)) converts the gravitational potential energy of water behind a dam to electric energy. (a) What is the gravitational potential energy relative to the generators of a lake of volume 50.0 km^3 (mass $= 5.00 \times 10^{13} \text{ kg}$), given that the lake has an average height of 40.0 m above the generators? (b) Compare this with the energy stored in a 9-megaton fusion bomb.



Hydroelectric facility (credit: Denis

Solution:

(a) $1.96 \times 10^{16} \text{ J}$

(b) The ratio of gravitational potential energy in the lake to the energy stored in the bomb is 0.52. That is, the energy stored in the lake is approximately half that in a 9-megaton fusion bomb.

Exercise:

Problem:

(a) How much gravitational potential energy (relative to the ground on which it is built) is stored in the Great Pyramid of Cheops, given that its mass is about $7 \times 10^9 \text{ kg}$ and its center of mass is 36.5 m above the surrounding ground? (b) How does this energy compare with the daily food intake of a person?

Exercise:

Problem:

Suppose a 350-g kookaburra (a large kingfisher bird) picks up a 75-g snake and raises it 2.5 m from the ground to a branch. (a) How much work did the bird do on the snake? (b) How much work did it do to raise its own center of mass to the branch?

Solution:

(a) 1.8 J

(b) 8.6 J

Exercise:

Problem:

In [\[link\]](#), we found that the speed of a roller coaster that had descended 20.0 m was only slightly greater when it had an initial speed of 5.00 m/s than when it started from rest. This implies that $\Delta PE \gg KE_i$. Confirm this statement by taking the ratio of ΔPE to KE_i . (Note that mass cancels.)

Exercise:

Problem:

A 100-g toy car is propelled by a compressed spring that starts it moving. The car follows the curved track in [\[link\]](#). Show that the final speed of the toy car is 0.687 m/s if its initial speed is 2.00 m/s and it coasts up the frictionless slope, gaining 0.180 m in altitude.



A toy car moves up a sloped track.
(credit: Leszek Leszczynski, Flickr)

Solution:**Equation:**

$$v_f = \sqrt{2gh + v_0^2} = \sqrt{2(9.80 \text{ m/s}^2)(-0.180 \text{ m}) + (2.00 \text{ m/s})^2} = 0.687 \text{ m/s}$$

Exercise:**Problem:**

In a downhill ski race, surprisingly, little advantage is gained by getting a running start. (This is because the initial kinetic energy is small compared with the gain in gravitational potential energy on even small hills.) To demonstrate this, find the final speed and the time taken for a skier who skies 70.0 m along a 30° slope neglecting friction: (a) Starting from rest. (b) Starting with an initial speed of 2.50 m/s. (c) Does the answer surprise you? Discuss why it is still advantageous to get a running start in very competitive events.

Glossary

gravitational potential energy

the energy an object has due to its position in a gravitational field

Conservative Forces and Potential Energy

- Define conservative force, potential energy, and mechanical energy.
- Explain the potential energy of a spring in terms of its compression when Hooke's law applies.
- Use the work-energy theorem to show how having only conservative forces implies conservation of mechanical energy.

Potential Energy and Conservative Forces

Work is done by a force, and some forces, such as weight, have special characteristics. A **conservative force** is one, like the gravitational force, for which work done by or against it depends only on the starting and ending points of a motion and not on the path taken. We can define a **potential energy** (PE) for any conservative force, just as we did for the gravitational force. For example, when you wind up a toy, an egg timer, or an old-fashioned watch, you do work against its spring and store energy in it. (We treat these springs as ideal, in that we assume there is no friction and no production of thermal energy.) This stored energy is recoverable as work, and it is useful to think of it as potential energy contained in the spring. Indeed, the reason that the spring has this characteristic is that its force is *conservative*. That is, a conservative force results in stored or potential energy. Gravitational potential energy is one example, as is the energy stored in a spring. We will also see how conservative forces are related to the conservation of energy.

Note:

Potential Energy and Conservative Forces

Potential energy is the energy a system has due to position, shape, or configuration. It is stored energy that is completely recoverable.

A conservative force is one for which work done by or against it depends only on the starting and ending points of a motion and not on the path taken.

We can define a potential energy (PE) for any conservative force. The work done against a conservative force to reach a final configuration

depends on the configuration, not the path followed, and is the potential energy added.

Potential Energy of a Spring

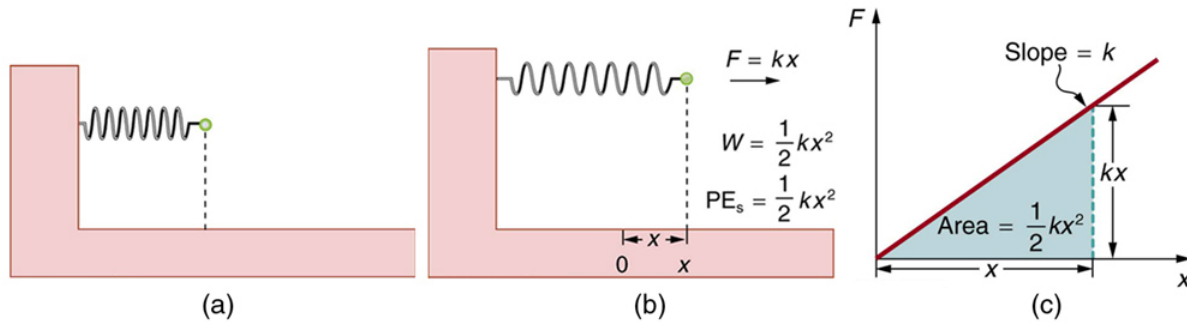
First, let us obtain an expression for the potential energy stored in a spring (PE_s). We calculate the work done to stretch or compress a spring that obeys Hooke's law. (Hooke's law was examined in [Elasticity: Stress and Strain](#), and states that the magnitude of force F on the spring and the resulting deformation ΔL are proportional, $F = k\Delta L$.) (See [\[link\]](#).) For our spring, we will replace ΔL (the amount of deformation produced by a force F) by the distance x that the spring is stretched or compressed along its length. So the force needed to stretch the spring has magnitude $F = kx$, where k is the spring's force constant. The force increases linearly from 0 at the start to kx in the fully stretched position. The average force is $kx/2$. Thus the work done in stretching or compressing the spring is

$W_s = Fd = \left(\frac{kx}{2}\right)x = \frac{1}{2}kx^2$. Alternatively, we noted in [Kinetic Energy and the Work-Energy Theorem](#) that the area under a graph of F vs. x is the work done by the force. In [\[link\]](#)(c) we see that this area is also $\frac{1}{2}kx^2$. We therefore define the **potential energy of a spring**, PE_s , to be

Equation:

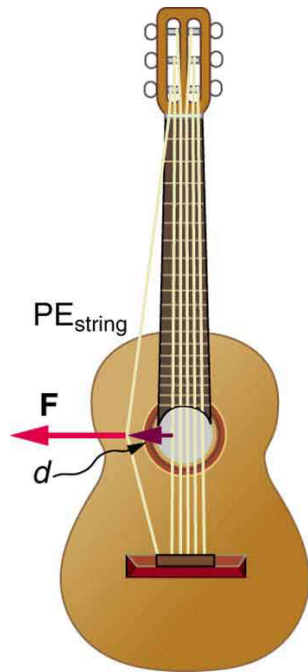
$$PE_s = \frac{1}{2}kx^2,$$

where k is the spring's force constant and x is the displacement from its undeformed position. The potential energy represents the work done *on* the spring and the energy stored in it as a result of stretching or compressing it a distance x . The potential energy of the spring PE_s does not depend on the path taken; it depends only on the stretch or squeeze x in the final configuration.



- (a) An undeformed spring has no PE_s stored in it. (b) The force needed to stretch (or compress) the spring a distance x has a magnitude $F = kx$, and the work done to stretch (or compress) it is $\frac{1}{2}kx^2$. Because the force is conservative, this work is stored as potential energy (PE_s) in the spring, and it can be fully recovered. (c) A graph of F vs. x has a slope of k , and the area under the graph is $\frac{1}{2}kx^2$. Thus the work done or potential energy stored is $\frac{1}{2}kx^2$.

The equation $PE_s = \frac{1}{2}kx^2$ has general validity beyond the special case for which it was derived. Potential energy can be stored in any elastic medium by deforming it. Indeed, the general definition of **potential energy** is energy due to position, shape, or configuration. For shape or position deformations, stored energy is $PE_s = \frac{1}{2}kx^2$, where k is the force constant of the particular system and x is its deformation. Another example is seen in [\[link\]](#) for a guitar string.



Work is done
to deform the
guitar string,
giving it
potential
energy.

When
released, the
potential
energy is
converted to
kinetic
energy and
back to
potential as
the string
oscillates
back and
forth. A very
small
fraction is
dissipated as

sound
energy,
slowly
removing
energy from
the string.

Conservation of Mechanical Energy

Let us now consider what form the work-energy theorem takes when only conservative forces are involved. This will lead us to the conservation of energy principle. The work-energy theorem states that the net work done by all forces acting on a system equals its change in kinetic energy. In equation form, this is

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2 = \Delta\text{KE}.$$

If only conservative forces act, then

Equation:

$$W_{\text{net}} = W_{\text{c}},$$

where W_{c} is the total work done by all conservative forces. Thus,

Equation:

$$W_{\text{c}} = \Delta\text{KE}.$$

Now, if the conservative force, such as the gravitational force or a spring force, does work, the system loses potential energy. That is, $W_{\text{c}} = -\Delta\text{PE}$. Therefore,

Equation:

$$-\Delta PE = \Delta KE$$

or

Equation:

$$\Delta KE + \Delta PE = 0.$$

This equation means that the total kinetic and potential energy is constant for any process involving only conservative forces. That is,

Equation:

$$KE + PE = \text{constant}$$

or

(conservative forces only),

$$KE_i + PE_i = KE_f + PE_f$$

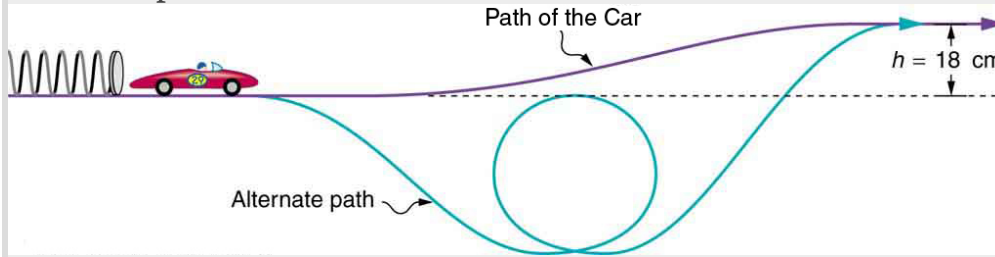
where i and f denote initial and final values. This equation is a form of the work-energy theorem for conservative forces; it is known as the **conservation of mechanical energy** principle. Remember that this applies to the extent that all the forces are conservative, so that friction is negligible. The total kinetic plus potential energy of a system is defined to be its **mechanical energy**, $(KE + PE)$. In a system that experiences only conservative forces, there is a potential energy associated with each force, and the energy only changes form between KE and the various types of PE, with the total energy remaining constant.

Example:

Using Conservation of Mechanical Energy to Calculate the Speed of a Toy Car

A 0.100-kg toy car is propelled by a compressed spring, as shown in [\[link\]](#). The car follows a track that rises 0.180 m above the starting point. The spring is compressed 4.00 cm and has a force constant of 250.0 N/m. Assuming work done by friction to be negligible, find (a) how fast the car

is going before it starts up the slope and (b) how fast it is going at the top of the slope.



A toy car is pushed by a compressed spring and coasts up a slope. Assuming negligible friction, the potential energy in the spring is first completely converted to kinetic energy, and then to a combination of kinetic and gravitational potential energy as the car rises. The details of the path are unimportant because all forces are conservative—the car would have the same final speed if it took the alternate path shown.

Strategy

The spring force and the gravitational force are conservative forces, so conservation of mechanical energy can be used. Thus,

Equation:

$$KE_i + PE_i = KE_f + PE_f$$

or

Equation:

$$\frac{1}{2}mv_i^2 + mgh_i + \frac{1}{2}kx_i^2 = \frac{1}{2}mv_f^2 + mgh_f + \frac{1}{2}kx_f^2,$$

where h is the height (vertical position) and x is the compression of the spring. This general statement looks complex but becomes much simpler when we start considering specific situations. First, we must identify the initial and final conditions in a problem; then, we enter them into the last equation to solve for an unknown.

Solution for (a)

This part of the problem is limited to conditions just before the car is released and just after it leaves the spring. Take the initial height to be zero, so that both h_i and h_f are zero. Furthermore, the initial speed v_i is zero and the final compression of the spring x_f is zero, and so several terms in the conservation of mechanical energy equation are zero and it simplifies to

Equation:

$$\frac{1}{2} k x_i^2 = \frac{1}{2} m v_f^2.$$

In other words, the initial potential energy in the spring is converted completely to kinetic energy in the absence of friction. Solving for the final speed and entering known values yields

Equation:

$$\begin{aligned} v_f &= \sqrt{\frac{k}{m}} x_i \\ &= \sqrt{\frac{250.0 \text{ N/m}}{0.100 \text{ kg}}} (0.0400 \text{ m}) \\ &= 2.00 \text{ m/s.} \end{aligned}$$

Solution for (b)

One method of finding the speed at the top of the slope is to consider conditions just before the car is released and just after it reaches the top of the slope, completely ignoring everything in between. Doing the same type of analysis to find which terms are zero, the conservation of mechanical energy becomes

Equation:

$$\frac{1}{2} k x_i^2 = \frac{1}{2} m v_f^2 + m g h_f.$$

This form of the equation means that the spring's initial potential energy is converted partly to gravitational potential energy and partly to kinetic energy. The final speed at the top of the slope will be less than at the bottom. Solving for v_f and substituting known values gives

Equation:

$$\begin{aligned}
 v_f &= \sqrt{\frac{kx_i^2}{m} - 2gh_f} \\
 &= \sqrt{\left(\frac{250.0 \text{ N/m}}{0.100 \text{ kg}}\right)(0.0400 \text{ m})^2 - 2(9.80 \text{ m/s}^2)(0.180 \text{ m})} \\
 &= 0.687 \text{ m/s.}
 \end{aligned}$$

Discussion

Another way to solve this problem is to realize that the car's kinetic energy before it goes up the slope is converted partly to potential energy—that is, to take the final conditions in part (a) to be the initial conditions in part (b).

Note that, for conservative forces, we do not directly calculate the work they do; rather, we consider their effects through their corresponding potential energies, just as we did in [\[link\]](#). Note also that we do not consider details of the path taken—only the starting and ending points are important (as long as the path is not impossible). This assumption is usually a tremendous simplification, because the path may be complicated and forces may vary along the way.

Note:

PhET Explorations: Energy Skate Park

Learn about conservation of energy with a skater dude! Build tracks, ramps and jumps for the skater and view the kinetic energy, potential energy and friction as he moves. You can also take the skater to different planets or even space!

https://phet.colorado.edu/sims/html/energy-skate-park-basics/latest/energy-skate-park-basics_en.html

Section Summary

- A conservative force is one for which work depends only on the starting and ending points of a motion, not on the path taken.
- We can define potential energy (PE) for any conservative force, just as we defined PE_g for the gravitational force.
- The potential energy of a spring is $PE_s = \frac{1}{2}kx^2$, where k is the spring's force constant and x is the displacement from its undeformed position.
- Mechanical energy is defined to be $KE + PE$ for a conservative force.
- When only conservative forces act on and within a system, the total mechanical energy is constant. In equation form,

Equation:

$$KE + PE = \text{constant}$$

or

$$KE_i + PE_i = KE_f + PE_f$$

where i and f denote initial and final values. This is known as the conservation of mechanical energy.

Conceptual Questions

Exercise:

Problem: What is a conservative force?

Exercise:

Problem:

The force exerted by a diving board is conservative, provided the internal friction is negligible. Assuming friction is negligible, describe changes in the potential energy of a diving board as a swimmer dives from it, starting just before the swimmer steps on the board until just after his feet leave it.

Exercise:

Problem:

Define mechanical energy. What is the relationship of mechanical energy to nonconservative forces? What happens to mechanical energy if only conservative forces act?

Exercise:**Problem:**

What is the relationship of potential energy to conservative force?

Problems & Exercises**Exercise:****Problem:**

A 5.00×10^5 -kg subway train is brought to a stop from a speed of 0.500 m/s in 0.400 m by a large spring bumper at the end of its track. What is the force constant k of the spring?

Solution:**Equation:**

$$7.81 \times 10^5 \text{ N/m}$$

Exercise:**Problem:**

A pogo stick has a spring with a force constant of 2.50×10^4 N/m, which can be compressed 12.0 cm. To what maximum height can a child jump on the stick using only the energy in the spring, if the child and stick have a total mass of 40.0 kg? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Energy](#).

Glossary

conservative force

a force that does the same work for any given initial and final configuration, regardless of the path followed

potential energy

energy due to position, shape, or configuration

potential energy of a spring

the stored energy of a spring as a function of its displacement; when Hooke's law applies, it is given by the expression $\frac{1}{2}kx^2$ where x is the distance the spring is compressed or extended and k is the spring constant

conservation of mechanical energy

the rule that the sum of the kinetic energies and potential energies remains constant if only conservative forces act on and within a system

mechanical energy

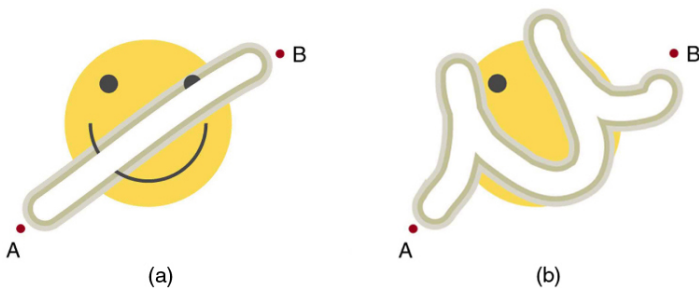
the sum of kinetic energy and potential energy

Nonconservative Forces

- Define nonconservative forces and explain how they affect mechanical energy.
- Show how the principle of conservation of energy can be applied by treating the conservative forces in terms of their potential energies and any nonconservative forces in terms of the work they do.

Nonconservative Forces and Friction

Forces are either conservative or nonconservative. Conservative forces were discussed in [Conservative Forces and Potential Energy](#). A **nonconservative force** is one for which work depends on the path taken. Friction is a good example of a nonconservative force. As illustrated in [\[link\]](#), work done against friction depends on the length of the path between the starting and ending points. Because of this dependence on path, there is no potential energy associated with nonconservative forces. An important characteristic is that the work done by a nonconservative force *adds or removes mechanical energy from a system*. **Friction**, for example, creates **thermal energy** that dissipates, removing energy from the system. Furthermore, even if the thermal energy is retained or captured, it cannot be fully converted back to work, so it is lost or not recoverable in that sense as well.

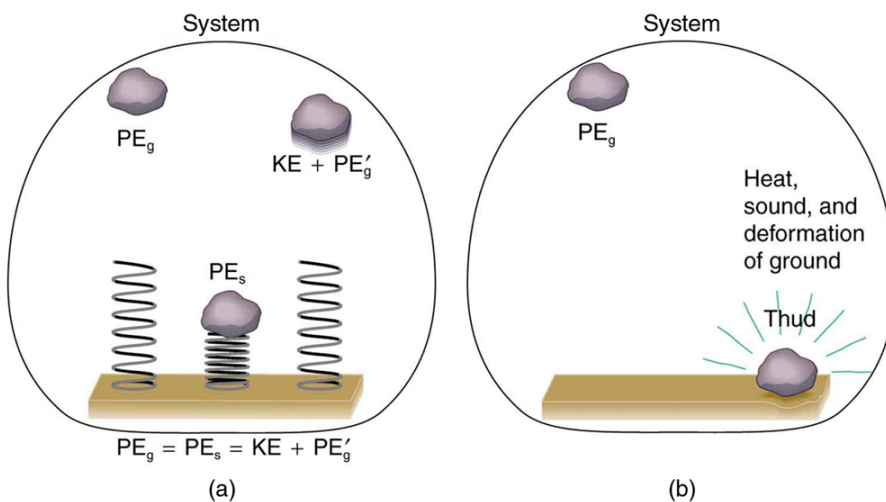


The amount of the happy face erased depends on the path taken by the eraser between points A and B, as does the work done against friction. Less work is done and less of the face

is erased for the path in (a) than for the path in (b). The force here is friction, and most of the work goes into thermal energy that subsequently leaves the system (the happy face plus the eraser). The energy expended cannot be fully recovered.

How Nonconservative Forces Affect Mechanical Energy

Mechanical energy may not be conserved when nonconservative forces act. For example, when a car is brought to a stop by friction on level ground, it loses kinetic energy, which is dissipated as thermal energy, reducing its mechanical energy. [\[link\]](#) compares the effects of conservative and nonconservative forces. We often choose to understand simpler systems such as that described in [\[link\]](#)(a) first before studying more complicated systems as in [\[link\]](#)(b).



Comparison of the effects of conservative and nonconservative forces on the mechanical energy of a system. (a) A system with only conservative

forces. When a rock is dropped onto a spring, its mechanical energy remains constant (neglecting air resistance) because the force in the spring is conservative. The spring can propel the rock back to its original height, where it once again has only potential energy due to gravity. (b) A system with nonconservative forces. When the same rock is dropped onto the ground, it is stopped by nonconservative forces that dissipate its mechanical energy as thermal energy, sound, and surface distortion. The rock has lost mechanical energy.

How the Work-Energy Theorem Applies

Now let us consider what form the work-energy theorem takes when both conservative and nonconservative forces act. We will see that the work done by nonconservative forces equals the change in the mechanical energy of a system. As noted in [Kinetic Energy and the Work-Energy Theorem](#), the work-energy theorem states that the net work on a system equals the change in its kinetic energy, or $W_{\text{net}} = \Delta\text{KE}$. The net work is the sum of the work by nonconservative forces plus the work by conservative forces. That is,

Equation:

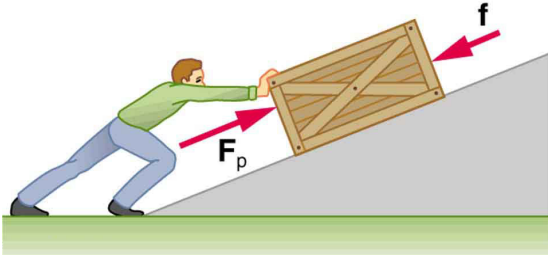
$$W_{\text{net}} = W_{\text{nc}} + W_{\text{c}},$$

so that

Equation:

$$W_{\text{nc}} + W_{\text{c}} = \Delta\text{KE},$$

where W_{nc} is the total work done by all nonconservative forces and W_{c} is the total work done by all conservative forces.



A person pushes a crate up a ramp, doing work on the crate. Friction and gravitational force (not shown) also do work on the crate; both forces oppose the person's push. As the crate is pushed up the ramp, it gains mechanical energy, implying that the work done by the person is greater than the work done by friction.

Consider [\[link\]](#), in which a person pushes a crate up a ramp and is opposed by friction. As in the previous section, we note that work done by a conservative force comes from a loss of gravitational potential energy, so that $W_c = -\Delta PE$. Substituting this equation into the previous one and solving for W_{nc} gives

Equation:

$$W_{nc} = \Delta KE + \Delta PE.$$

This equation means that the total mechanical energy ($KE + PE$) changes by exactly the amount of work done by nonconservative forces. In [\[link\]](#), this is the work done by the person minus the work done by friction. So even if energy is not conserved for the system of interest (such as the crate), we know that an equal amount of work was done to cause the change in total mechanical energy.

We rearrange $W_{\text{nc}} = \Delta\text{KE} + \Delta\text{PE}$ to obtain

Equation:

$$\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f.$$

This means that the amount of work done by nonconservative forces adds to the mechanical energy of a system. If W_{nc} is positive, then mechanical energy is increased, such as when the person pushes the crate up the ramp in [\[link\]](#). If W_{nc} is negative, then mechanical energy is decreased, such as when the rock hits the ground in [\[link\]](#)(b). If W_{nc} is zero, then mechanical energy is conserved, and nonconservative forces are balanced. For example, when you push a lawn mower at constant speed on level ground, your work done is removed by the work of friction, and the mower has a constant energy.

Applying Energy Conservation with Nonconservative Forces

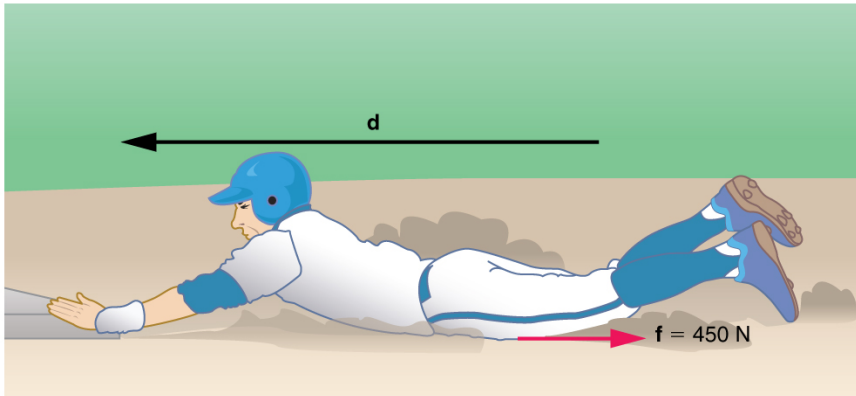
When no change in potential energy occurs, applying $\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f$ amounts to applying the work-energy theorem by setting the change in kinetic energy to be equal to the net work done on the system, which in the most general case includes both conservative and nonconservative forces. But when seeking instead to find a change in total mechanical energy in situations that involve changes in both potential and kinetic energy, the previous equation $\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f$ says that you can start by finding the change in mechanical energy that would have resulted from just the conservative forces, including the potential energy changes, and add to it the work done, with the proper sign, by any nonconservative forces involved.

Example:

Calculating Distance Traveled: How Far a Baseball Player Slides

Consider the situation shown in [\[link\]](#), where a baseball player slides to a stop on level ground. Using energy considerations, calculate the distance

the 65.0-kg baseball player slides, given that his initial speed is 6.00 m/s and the force of friction against him is a constant 450 N.



The baseball player slides to a stop in a distance d . In the process, friction removes the player's kinetic energy by doing an amount of work fd equal to the initial kinetic energy.

Strategy

Friction stops the player by converting his kinetic energy into other forms, including thermal energy. In terms of the work-energy theorem, the work done by friction, which is negative, is added to the initial kinetic energy to reduce it to zero. The work done by friction is negative, because \mathbf{f} is in the opposite direction of the motion (that is, $\theta = 180^\circ$, and so $\cos \theta = -1$). Thus $W_{\text{nc}} = -fd$. The equation simplifies to

Equation:

$$\frac{1}{2}mv_i^2 - fd = 0$$

or

Equation:

$$fd = \frac{1}{2}mv_i^2.$$

This equation can now be solved for the distance d .

Solution

Solving the previous equation for d and substituting known values yields
Equation:

$$\begin{aligned}d &= \frac{mv_i^2}{2f} \\&= \frac{(65.0 \text{ kg})(6.00 \text{ m/s})^2}{(2)(450 \text{ N})} \\&= 2.60 \text{ m.}\end{aligned}$$

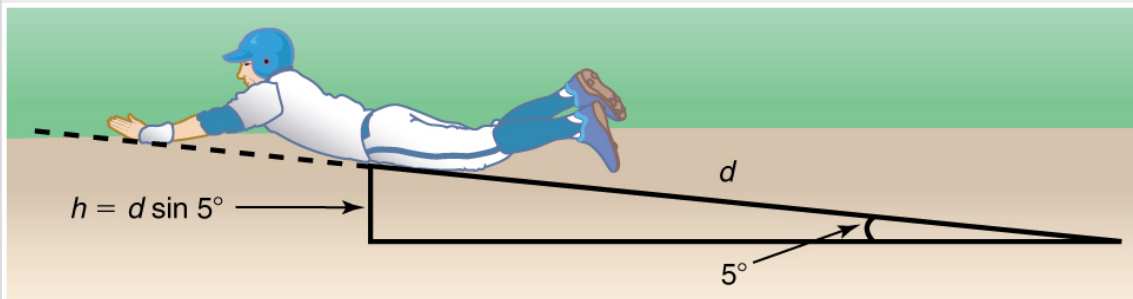
Discussion

The most important point of this example is that the amount of nonconservative work equals the change in mechanical energy. For example, you must work harder to stop a truck, with its large mechanical energy, than to stop a mosquito.

Example:

Calculating Distance Traveled: Sliding Up an Incline

Suppose that the player from [\[link\]](#) is running up a hill having a 5.00° incline upward with a surface similar to that in the baseball stadium. The player slides with the same initial speed, and the frictional force is still 450 N. Determine how far he slides.



The same baseball player slides to a stop on a 5.00° slope.

Strategy

In this case, the work done by the nonconservative friction force on the player reduces the mechanical energy he has from his kinetic energy at zero height, to the final mechanical energy he has by moving through

distance d to reach height h along the hill, with $h = d \sin 5.00^\circ$. This is expressed by the equation

Equation:

$$\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f.$$

Solution

The work done by friction is again $W_{\text{nc}} = -fd$; initially the potential energy is $\text{PE}_i = mg \cdot 0 = 0$ and the kinetic energy is $\text{KE}_i = \frac{1}{2}mv_i^2$; the final energy contributions are $\text{KE}_f = 0$ for the kinetic energy and $\text{PE}_f = mgh = mgd \sin \theta$ for the potential energy.

Substituting these values gives

Equation:

$$\frac{1}{2}mv_i^2 + 0 + (-fd) = 0 + mgd \sin \theta.$$

Solve this for d to obtain

Equation:

$$\begin{aligned} d &= \frac{(\frac{1}{2})mv_i^2}{f+mg \sin \theta} \\ &= \frac{(0.5)(65.0 \text{ kg})(6.00 \text{ m/s})^2}{450 \text{ N}+(65.0 \text{ kg})(9.80 \text{ m/s}^2) \sin (5.00^\circ)} \\ &= 2.31 \text{ m.} \end{aligned}$$

Discussion

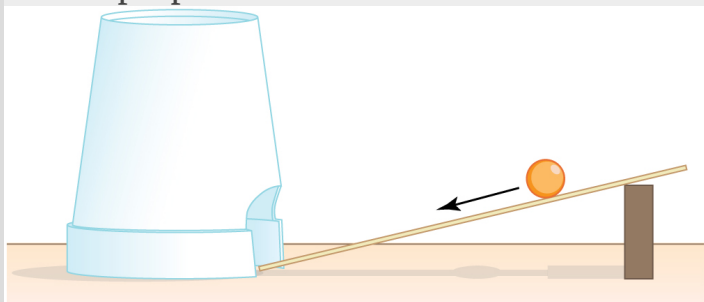
As might have been expected, the player slides a shorter distance by sliding uphill. Note that the problem could also have been solved in terms of the forces directly and the work energy theorem, instead of using the potential energy. This method would have required combining the normal force and force of gravity vectors, which no longer cancel each other because they point in different directions, and friction, to find the net force. You could then use the net force and the net work to find the distance d that reduces the kinetic energy to zero. By applying conservation of energy and using the potential energy instead, we need only consider the gravitational potential energy mgh , without combining and resolving force vectors. This simplifies the solution considerably.

Note:**Making Connections: Take-Home Investigation—Determining Friction from the Stopping Distance**

This experiment involves the conversion of gravitational potential energy into thermal energy. Use the ruler, book, and marble from [Take-Home Investigation—Converting Potential to Kinetic Energy](#). In addition, you will need a foam cup with a small hole in the side, as shown in [\[link\]](#). From the 10-cm position on the ruler, let the marble roll into the cup positioned at the bottom of the ruler. Measure the distance d the cup moves before stopping. What forces caused it to stop? What happened to the kinetic energy of the marble at the bottom of the ruler? Next, place the marble at the 20-cm and the 30-cm positions and again measure the distance the cup moves after the marble enters it. Plot the distance the cup moves versus the initial marble position on the ruler. Is this relationship linear?

With some simple assumptions, you can use these data to find the coefficient of kinetic friction μ_k of the cup on the table. The force of friction f on the cup is $\mu_k N$, where the normal force N is just the weight of the cup plus the marble. The normal force and force of gravity do no work because they are perpendicular to the displacement of the cup, which moves horizontally. The work done by friction is fd . You will need the mass of the marble as well to calculate its initial kinetic energy.

It is interesting to do the above experiment also with a steel marble (or ball bearing). Releasing it from the same positions on the ruler as you did with the glass marble, is the velocity of this steel marble the same as the velocity of the marble at the bottom of the ruler? Is the distance the cup moves proportional to the mass of the steel and glass marbles?



Rolling a marble down a ruler into a foam cup.

Note:**PhET Explorations: The Ramp**

Explore forces, energy and work as you push household objects up and down a ramp. Lower and raise the ramp to see how the angle of inclination affects the parallel forces acting on the file cabinet. Graphs show forces, energy and work.

[The
Ramp](#)
p

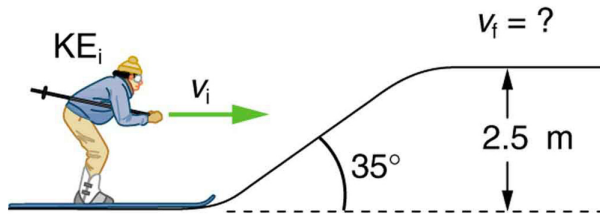
Section Summary

- A nonconservative force is one for which work depends on the path.
- Friction is an example of a nonconservative force that changes mechanical energy into thermal energy.
- Work W_{nc} done by a nonconservative force changes the mechanical energy of a system. In equation form, $W_{nc} = \Delta KE + \Delta PE$ or, equivalently, $KE_i + PE_i + W_{nc} = KE_f + PE_f$.
- When both conservative and nonconservative forces act, energy conservation can be applied and used to calculate motion in terms of the known potential energies of the conservative forces and the work done by nonconservative forces, instead of finding the net work from the net force, or having to directly apply Newton's laws.

Problems & Exercises**Exercise:**

Problem:

A 60.0-kg skier with an initial speed of 12.0 m/s coasts up a 2.50-m-high rise as shown in [\[link\]](#). Find her final speed at the top, given that the coefficient of friction between her skis and the snow is 0.0800. (Hint: Find the distance traveled up the incline assuming a straight-line path as shown in the figure.)



The skier's initial kinetic energy is partially used in coasting to the top of a rise.

Solution:

9.46 m/s

Exercise:**Problem:**

(a) How high a hill can a car coast up (engine disengaged) if work done by friction is negligible and its initial speed is 110 km/h? (b) If, in actuality, a 750-kg car with an initial speed of 110 km/h is observed to coast up a hill to a height 22.0 m above its starting point, how much thermal energy was generated by friction? (c) What is the average force of friction if the hill has a slope 2.5° above the horizontal?

Glossary

nonconservative force

a force whose work depends on the path followed between the given initial and final configurations

friction

the force between surfaces that opposes one sliding on the other;
friction changes mechanical energy into thermal energy

Conservation of Energy

- Explain the law of the conservation of energy.
- Describe some of the many forms of energy.
- Define efficiency of an energy conversion process as the fraction left as useful energy or work, rather than being transformed, for example, into thermal energy.

Law of Conservation of Energy

Energy, as we have noted, is conserved, making it one of the most important physical quantities in nature. The **law of conservation of energy** can be stated as follows:

Total energy is constant in any process. It may change in form or be transferred from one system to another, but the total remains the same.

We have explored some forms of energy and some ways it can be transferred from one system to another. This exploration led to the definition of two major types of energy—mechanical energy ($KE + PE$) and energy transferred via work done by nonconservative forces (W_{nc}). But energy takes *many* other forms, manifesting itself in *many* different ways, and we need to be able to deal with all of these before we can write an equation for the above general statement of the conservation of energy.

Other Forms of Energy than Mechanical Energy

At this point, we deal with all other forms of energy by lumping them into a single group called other energy (OE). Then we can state the conservation of energy in equation form as

Equation:

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f.$$

All types of energy and work can be included in this very general statement of conservation of energy. Kinetic energy is KE , work done by a conservative force is represented by PE , work done by nonconservative forces is W_{nc} , and

all other energies are included as OE. This equation applies to all previous examples; in those situations OE was constant, and so it subtracted out and was not directly considered.

Note:

Making Connections: Usefulness of the Energy Conservation Principle

The fact that energy is conserved and has many forms makes it very important. You will find that energy is discussed in many contexts, because it is involved in all processes. It will also become apparent that many situations are best understood in terms of energy and that problems are often most easily conceptualized and solved by considering energy.

When does OE play a role? One example occurs when a person eats. Food is oxidized with the release of carbon dioxide, water, and energy. Some of this chemical energy is converted to kinetic energy when the person moves, to potential energy when the person changes altitude, and to thermal energy (another form of OE).

Some of the Many Forms of Energy

What are some other forms of energy? You can probably name a number of forms of energy not yet discussed. Many of these will be covered in later chapters, but let us detail a few here. **Electrical energy** is a common form that is converted to many other forms and does work in a wide range of practical situations. Fuels, such as gasoline and food, carry **chemical energy** that can be transferred to a system through oxidation. Chemical fuel can also produce electrical energy, such as in batteries. Batteries can in turn produce light, which is a very pure form of energy. Most energy sources on Earth are in fact stored energy from the energy we receive from the Sun. We sometimes refer to this as **radiant energy**, or electromagnetic radiation, which includes visible light, infrared, and ultraviolet radiation. **Nuclear energy** comes from processes that convert measurable amounts of mass into energy. Nuclear energy is transformed into the energy of sunlight, into electrical energy in power plants, and into the energy of the heat transfer and blast in weapons.

Atoms and molecules inside all objects are in random motion. This internal mechanical energy from the random motions is called **thermal energy**, because it is related to the temperature of the object. These and all other forms of energy can be converted into one another and can do work.

[\[link\]](#) gives the amount of energy stored, used, or released from various objects and in various phenomena. The range of energies and the variety of types and situations is impressive.

Note:

Problem-Solving Strategies for Energy

You will find the following problem-solving strategies useful whenever you deal with energy. The strategies help in organizing and reinforcing energy concepts. In fact, they are used in the examples presented in this chapter. The familiar general problem-solving strategies presented earlier—involving identifying physical principles, knowns, and unknowns, checking units, and so on—continue to be relevant here.

Step 1. Determine the system of interest and identify what information is given and what quantity is to be calculated. A sketch will help.

Step 2. Examine all the forces involved and determine whether you know or are given the potential energy from the work done by the forces. Then use step 3 or step 4.

Step 3. If you know the potential energies for the forces that enter into the problem, then forces are all conservative, and you can apply conservation of mechanical energy simply in terms of potential and kinetic energy. The equation expressing conservation of energy is

Equation:

$$KE_i + PE_i = KE_f + PE_f.$$

Step 4. If you know the potential energy for only some of the forces, possibly because some of them are nonconservative and do not have a potential energy, or if there are other energies that are not easily treated in terms of force and work, then the conservation of energy law in its most general form must be used.

Equation:

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f.$$

In most problems, one or more of the terms is zero, simplifying its solution. Do not calculate W_c , the work done by conservative forces; it is already incorporated in the PE terms.

Step 5. You have already identified the types of work and energy involved (in step 2). Before solving for the unknown, *eliminate terms wherever possible* to simplify the algebra. For example, choose $h = 0$ at either the initial or final point, so that PE_g is zero there. Then solve for the unknown in the customary manner.

Step 6. *Check the answer to see if it is reasonable.* Once you have solved a problem, reexamine the forms of work and energy to see if you have set up the conservation of energy equation correctly. For example, work done against friction should be negative, potential energy at the bottom of a hill should be less than that at the top, and so on. Also check to see that the numerical value obtained is reasonable. For example, the final speed of a skateboarder who coasts down a 3-m-high ramp could reasonably be 20 km/h, but *not* 80 km/h.

Transformation of Energy

The transformation of energy from one form into others is happening all the time. The chemical energy in food is converted into thermal energy through metabolism; light energy is converted into chemical energy through photosynthesis. In a larger example, the chemical energy contained in coal is converted into thermal energy as it burns to turn water into steam in a boiler. This thermal energy in the steam in turn is converted to mechanical energy as it spins a turbine, which is connected to a generator to produce electrical energy. (In all of these examples, not all of the initial energy is converted into the forms mentioned. This important point is discussed later in this section.)

Another example of energy conversion occurs in a solar cell. Sunlight impinging on a solar cell (see [\[link\]](#)) produces electricity, which in turn can be used to run an electric motor. Energy is converted from the primary source of solar energy into electrical energy and then into mechanical energy.



Solar energy is converted into electrical energy by solar cells, which is used to run a motor in this solar-power aircraft. (credit: NASA)

Object/phenomenon	Energy in joules
Big Bang	10^{68}
Energy released in a supernova	10^{44}
Fusion of all the hydrogen in Earth's oceans	10^{34}
Annual world energy use	4×10^{20}

Object/phenomenon	Energy in joules
Large fusion bomb (9 megaton)	3.8×10^{16}
1 kg hydrogen (fusion to helium)	6.4×10^{14}
1 kg uranium (nuclear fission)	8.0×10^{13}
Hiroshima-size fission bomb (10 kiloton)	4.2×10^{13}
90,000-ton aircraft carrier at 30 knots	1.1×10^{10}
1 barrel crude oil	5.9×10^9
1 ton TNT	4.2×10^9
1 gallon of gasoline	1.2×10^8
Daily home electricity use (developed countries)	7×10^7
Daily adult food intake (recommended)	1.2×10^7

Object/phenomenon	Energy in joules
1000-kg car at 90 km/h	3.1×10^5
1 g fat (9.3 kcal)	3.9×10^4
ATP hydrolysis reaction	3.2×10^4
1 g carbohydrate (4.1 kcal)	1.7×10^4
1 g protein (4.1 kcal)	1.7×10^4
Tennis ball at 100 km/h	22
Mosquito (10^{-2} g at 0.5 m/s)	1.3×10^{-6}
Single electron in a TV tube beam	4.0×10^{-15}
Energy to break one DNA strand	10^{-19}

Energy of Various Objects and Phenomena

Efficiency

Even though energy is conserved in an energy conversion process, the output of *useful energy* or work will be less than the energy input. The **efficiency** Eff of an energy conversion process is defined as

Equation:

$$\text{Efficiency}(\text{Eff}) = \frac{\text{useful energy or work output}}{\text{total energy input}} = \frac{W_{\text{out}}}{E_{\text{in}}}.$$

[\[link\]](#) lists some efficiencies of mechanical devices and human activities. In a coal-fired power plant, for example, about 40% of the chemical energy in the coal becomes useful electrical energy. The other 60% transforms into other (perhaps less useful) energy forms, such as thermal energy, which is then released to the environment through combustion gases and cooling towers.

Activity/device	Efficiency (%) [footnote] Representative values
Cycling and climbing	20
Swimming, surface	2
Swimming, submerged	4
Shoveling	3
Weightlifting	9
Steam engine	17
Gasoline engine	30

Activity/device	Efficiency (%) ^[footnote] Representative values
Diesel engine	35
Nuclear power plant	35
Coal power plant	42
Electric motor	98
Compact fluorescent light	20
Gas heater (residential)	90
Solar cell	10

Efficiency of the Human Body and Mechanical Devices

Note:

PhET Explorations: Masses and Springs

A realistic mass and spring laboratory. Hang masses from springs and adjust the spring stiffness and damping. You can even slow time. Transport the lab to different planets. A chart shows the kinetic, potential, and thermal energies for each spring.

https://phet.colorado.edu/sims/mass-spring-lab/mass-spring-lab_en.html

Section Summary

- The law of conservation of energy states that the total energy is constant in any process. Energy may change in form or be transferred from one system to another, but the total remains the same.
- When all forms of energy are considered, conservation of energy is written in equation form as

$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f$, where OE is all **other forms of energy** besides mechanical energy.

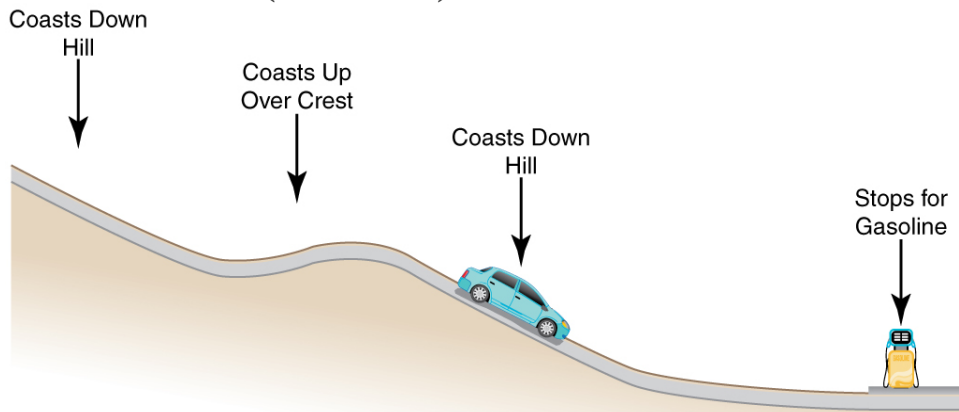
- Commonly encountered forms of energy include electric energy, chemical energy, radiant energy, nuclear energy, and thermal energy.
- Energy is often utilized to do work, but it is not possible to convert all the energy of a system to work.
- The efficiency Eff of a machine or human is defined to be $Eff = \frac{W_{out}}{E_{in}}$, where W_{out} is useful work output and E_{in} is the energy consumed.

Conceptual Questions

Exercise:

Problem:

Consider the following scenario. A car for which friction is *not* negligible accelerates from rest down a hill, running out of gasoline after a short distance. The driver lets the car coast farther down the hill, then up and over a small crest. He then coasts down that hill into a gas station, where he brakes to a stop and fills the tank with gasoline. Identify the forms of energy the car has, and how they are changed and transferred in this series of events. (See [\[link\]](#).)



A car experiencing non-negligible friction coasts down a hill, over a small crest, then downhill again, and comes to a stop at a gas station.

Exercise:**Problem:**

Describe the energy transfers and transformations for a javelin, starting from the point at which an athlete picks up the javelin and ending when the javelin is stuck into the ground after being thrown.

Exercise:**Problem:**

Do devices with efficiencies of less than one violate the law of conservation of energy? Explain.

Exercise:**Problem:**

List four different forms or types of energy. Give one example of a conversion from each of these forms to another form.

Exercise:

Problem: List the energy conversions that occur when riding a bicycle.

Problems & Exercises**Exercise:****Problem:**

Using values from [\[link\]](#), how many DNA molecules could be broken by the energy carried by a single electron in the beam of an old-fashioned TV tube? (These electrons were not dangerous in themselves, but they did create dangerous x rays. Later model tube TVs had shielding that absorbed x rays before they escaped and exposed viewers.)

Solution:

4×10^4 molecules

Exercise:**Problem:**

Using energy considerations and assuming negligible air resistance, show that a rock thrown from a bridge 20.0 m above water with an initial speed of 15.0 m/s strikes the water with a speed of 24.8 m/s independent of the direction thrown.

Solution:

Equating ΔPE_g and ΔKE , we obtain

$$v = \sqrt{2gh + v_0^2} = \sqrt{2(9.80 \text{ m/s}^2)(20.0 \text{ m}) + (15.0 \text{ m/s})^2} = 24.8 \text{ m/s}$$

Exercise:**Problem:**

If the energy in fusion bombs were used to supply the energy needs of the world, how many of the 9-megaton variety would be needed for a year's supply of energy (using data from [\[link\]](#))? This is not as far-fetched as it may sound—there are thousands of nuclear bombs, and their energy can be trapped in underground explosions and converted to electricity, as natural geothermal energy is.

Exercise:**Problem:**

(a) Use of hydrogen fusion to supply energy is a dream that may be realized in the next century. Fusion would be a relatively clean and almost limitless supply of energy, as can be seen from [\[link\]](#). To illustrate this, calculate how many years the present energy needs of the world could be supplied by one millionth of the oceans' hydrogen fusion energy. (b) How does this time compare with historically significant events, such as the duration of stable economic systems?

Solution:

(a) 25×10^6 years

(b) This is much, much longer than human time scales.

Glossary

law of conservation of energy

the general law that total energy is constant in any process; energy may change in form or be transferred from one system to another, but the total remains the same

electrical energy

the energy carried by a flow of charge

chemical energy

the energy in a substance stored in the bonds between atoms and molecules that can be released in a chemical reaction

radiant energy

the energy carried by electromagnetic waves

nuclear energy

energy released by changes within atomic nuclei, such as the fusion of two light nuclei or the fission of a heavy nucleus

thermal energy

the energy within an object due to the random motion of its atoms and molecules that accounts for the object's temperature

efficiency

a measure of the effectiveness of the input of energy to do work; useful energy or work divided by the total input of energy

Power

- Calculate power by calculating changes in energy over time.
- Examine power consumption and calculations of the cost of energy consumed.

What is Power?

Power—the word conjures up many images: a professional football player muscling aside his opponent, a dragster roaring away from the starting line, a volcano blowing its lava into the atmosphere, or a rocket blasting off, as in [\[link\]](#).



This powerful rocket on the Space Shuttle *Endeavor* did work and consumed energy at a very high rate. (credit: NASA)

These images of power have in common the rapid performance of work, consistent with the scientific definition of **power** (P) as the rate at which work is done.

Note:**Power**

Power is the rate at which work is done.

Equation:

$$P = \frac{W}{t}$$

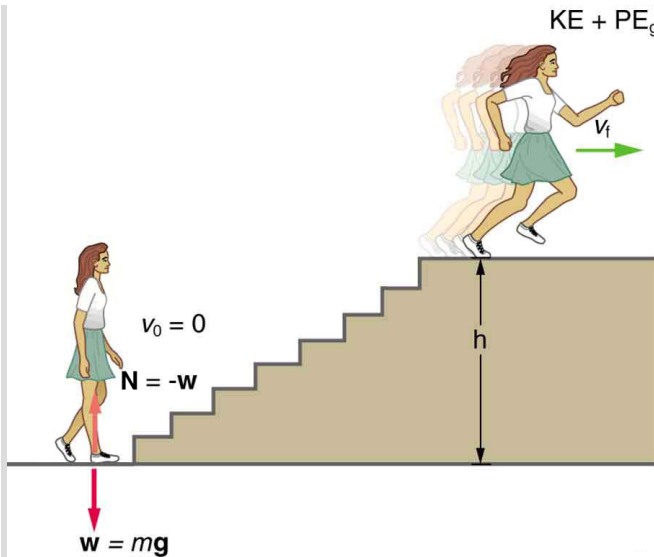
The SI unit for power is the **watt** (W), where 1 watt equals 1 joule/second ($1 \text{ W} = 1 \text{ J/s}$).

Because work is energy transfer, power is also the rate at which energy is expended. A 60-W light bulb, for example, expends 60 J of energy per second. Great power means a large amount of work or energy developed in a short time. For example, when a powerful car accelerates rapidly, it does a large amount of work and consumes a large amount of fuel in a short time.

Calculating Power from Energy

Example:**Calculating the Power to Climb Stairs**

What is the power output for a 60.0-kg woman who runs up a 3.00 m high flight of stairs in 3.50 s, starting from rest but having a final speed of 2.00 m/s? (See [\[link\]](#).)



When this woman runs upstairs starting from rest, she converts the chemical energy originally from food into kinetic energy and gravitational potential energy. Her power output depends on how fast she does this.

Strategy and Concept

The work going into mechanical energy is $W = KE + PE$. At the bottom of the stairs, we take both KE and PE_g as initially zero; thus,

$W = KE_f + PE_g = \frac{1}{2}mv_f^2 + mgh$, where h is the vertical height of the stairs. Because all terms are given, we can calculate W and then divide it by time to get power.

Solution

Substituting the expression for W into the definition of power given in the previous equation, $P = W/t$ yields

Equation:

$$P = \frac{W}{t} = \frac{\frac{1}{2}mv_f^2 + mgh}{t}.$$

Entering known values yields

Equation:

$$\begin{aligned} P &= \frac{0.5(60.0 \text{ kg})(2.00 \text{ m/s})^2 + (60.0 \text{ kg})(9.80 \text{ m/s}^2)(3.00 \text{ m})}{3.50 \text{ s}} \\ &= \frac{120 \text{ J} + 1764 \text{ J}}{3.50 \text{ s}} \\ &= 538 \text{ W}. \end{aligned}$$

Discussion

The woman does 1764 J of work to move up the stairs compared with only 120 J to increase her kinetic energy; thus, most of her power output is required for climbing rather than accelerating.

It is impressive that this woman's useful power output is slightly less than 1 **horsepower** (1 hp = 746 W)! People can generate more than a horsepower with their leg muscles for short periods of time by rapidly converting available blood sugar and oxygen into work output. (A horse can put out 1 hp for hours on end.) Once oxygen is depleted, power output decreases and the person begins to breathe rapidly to obtain oxygen to metabolize more food—this is known as the *aerobic* stage of exercise. If the woman climbed the stairs slowly, then her power output would be much less, although the amount of work done would be the same.

Note:**Making Connections: Take-Home Investigation—Measure Your Power Rating**

Determine your own power rating by measuring the time it takes you to climb a flight of stairs. We will ignore the gain in kinetic energy, as the above example showed that it was a small portion of the energy gain. Don't expect that your output will be more than about 0.5 hp.

Examples of Power

Examples of power are limited only by the imagination, because there are as many types as there are forms of work and energy. (See [\[link\]](#) for some examples.) Sunlight reaching Earth's surface carries a maximum power of about 1.3 kilowatts per square meter (kW/m^2). A tiny fraction of this is retained by Earth over the long term. Our consumption rate of fossil fuels is far greater than the rate at which they are stored, so it is inevitable that they will be depleted. Power implies that energy is transferred, perhaps changing form. It is never possible to change one form completely into another without losing some of it as thermal energy. For example, a 60-W incandescent bulb converts only 5 W of electrical power to light, with 55 W dissipating into thermal energy. Furthermore, the typical electric power plant converts only 35 to 40% of its fuel into electricity. The remainder becomes a huge amount of thermal energy that must be dispersed as heat transfer, as rapidly as it is created. A coal-fired power plant may produce 1000 megawatts; 1 megawatt (MW) is 10^6 W of electric power. But the power plant consumes chemical energy at a rate of about 2500 MW, creating heat transfer to the surroundings at a rate of 1500 MW. (See [\[link\]](#).)



Tremendous amounts of electric power are generated by coal-fired power plants such as this one in China, but an even larger amount of power goes into heat transfer to the surroundings.

The large cooling towers here are needed to transfer heat as rapidly as it is produced. The transfer of heat is not unique to coal plants but is an unavoidable consequence of generating electric power from any fuel—nuclear, coal, oil, natural gas, or the like. (credit: Kleinolive, Wikimedia Commons)

Object or Phenomenon	Power in Watts
Supernova (at peak)	5×10^{37}
Milky Way galaxy	10^{37}
Crab Nebula pulsar	10^{28}
The Sun	4×10^{26}

Object or Phenomenon	Power in Watts
Volcanic eruption (maximum)	4×10^{15}
Lightning bolt	2×10^{12}
Nuclear power plant (total electric and heat transfer)	3×10^9
Aircraft carrier (total useful and heat transfer)	10^8
Dragster (total useful and heat transfer)	2×10^6
Car (total useful and heat transfer)	8×10^4
Football player (total useful and heat transfer)	5×10^3
Clothes dryer	4×10^3
Person at rest (all heat transfer)	100

Object or Phenomenon	Power in Watts
Typical incandescent light bulb (total useful and heat transfer)	60
Heart, person at rest (total useful and heat transfer)	8
Electric clock	3
Pocket calculator	10^{-3}

Power Output or Consumption

Power and Energy Consumption

We usually have to pay for the energy we use. It is interesting and easy to estimate the cost of energy for an electrical appliance if its power consumption rate and time used are known. The higher the power consumption rate and the longer the appliance is used, the greater the cost of that appliance. The power consumption rate is $P = W/t = E/t$, where E is the energy supplied by the electricity company. So the energy consumed over a time t is

Equation:

$$E = Pt.$$

Electricity bills state the energy used in units of **kilowatt-hours** ($\text{kW} \cdot \text{h}$), which is the product of power in kilowatts and time in hours. This unit is convenient because electrical power consumption at the kilowatt level for hours at a time is typical.

Example:**Calculating Energy Costs**

What is the cost of running a 0.200-kW computer 6.00 h per day for 30.0 d if the cost of electricity is \$0.120 per kW · h?

Strategy

Cost is based on energy consumed; thus, we must find E from $E = Pt$ and then calculate the cost. Because electrical energy is expressed in kW · h, at the start of a problem such as this it is convenient to convert the units into kW and hours.

Solution

The energy consumed in kW · h is

Equation:

$$\begin{aligned} E &= Pt = (0.200 \text{ kW})(6.00 \text{ h/d})(30.0 \text{ d}) \\ &= 36.0 \text{ kW} \cdot \text{h}, \end{aligned}$$

and the cost is simply given by

Equation:

$$\text{cost} = (36.0 \text{ kW} \cdot \text{h})(\$0.120 \text{ per kW} \cdot \text{h}) = \$4.32 \text{ per month.}$$

Discussion

The cost of using the computer in this example is neither exorbitant nor negligible. It is clear that the cost is a combination of power and time. When both are high, such as for an air conditioner in the summer, the cost is high.

The motivation to save energy has become more compelling with its ever-increasing price. Armed with the knowledge that energy consumed is the product of power and time, you can estimate costs for yourself and make the necessary value judgments about where to save energy. Either power or time must be reduced. It is most cost-effective to limit the use of high-power devices that normally operate for long periods of time, such as water heaters and air conditioners. This would not include relatively high power devices like toasters, because they are on only a few minutes per day. It would also not include electric clocks, in spite of their 24-hour-per-day

usage, because they are very low power devices. It is sometimes possible to use devices that have greater efficiencies—that is, devices that consume less power to accomplish the same task. One example is the compact fluorescent light bulb, which produces over four times more light per watt of power consumed than its incandescent cousin.

Modern civilization depends on energy, but current levels of energy consumption and production are not sustainable. The likelihood of a link between global warming and fossil fuel use (with its concomitant production of carbon dioxide), has made reduction in energy use as well as a shift to non-fossil fuels of the utmost importance. Even though energy in an isolated system is a conserved quantity, the final result of most energy transformations is waste heat transfer to the environment, which is no longer useful for doing work. As we will discuss in more detail in [Thermodynamics](#), the potential for energy to produce useful work has been “degraded” in the energy transformation.

Section Summary

- Power is the rate at which work is done, or in equation form, for the average power P for work W done over a time t , $P = W/t$.
- The SI unit for power is the watt (W), where $1 \text{ W} = 1 \text{ J/s}$.
- The power of many devices such as electric motors is also often expressed in horsepower (hp), where $1 \text{ hp} = 746 \text{ W}$.

Conceptual Questions

Exercise:

Problem:

Most electrical appliances are rated in watts. Does this rating depend on how long the appliance is on? (When off, it is a zero-watt device.) Explain in terms of the definition of power.

Exercise:

Problem:

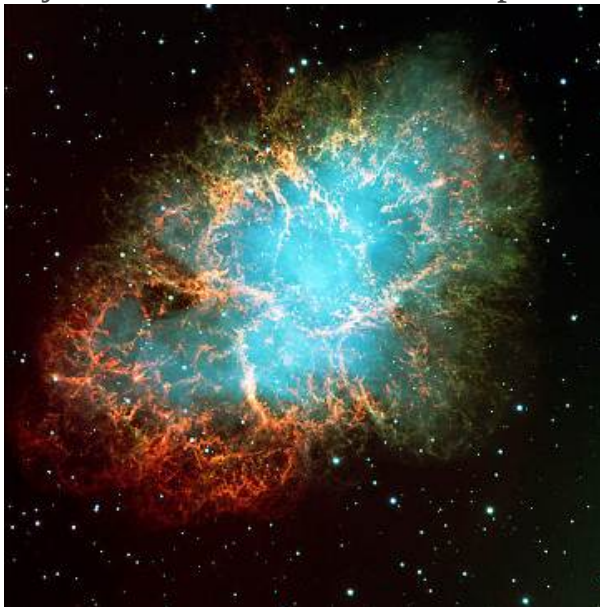
Explain, in terms of the definition of power, why energy consumption is sometimes listed in kilowatt-hours rather than joules. What is the relationship between these two energy units?

Exercise:**Problem:**

A spark of static electricity, such as that you might receive from a doorknob on a cold dry day, may carry a few hundred watts of power. Explain why you are not injured by such a spark.

Problems & Exercises**Exercise:****Problem:**

The Crab Nebula (see [\[link\]](#)) pulsar is the remnant of a supernova that occurred in A.D. 1054. Using data from [\[link\]](#), calculate the approximate factor by which the power output of this astronomical object has declined since its explosion.



Crab Nebula (credit: ESO, via
Wikimedia Commons)

Solution:

Equation:

$$2 \times 10^{-10}$$

Exercise:

Problem:

Suppose a star 1000 times brighter than our Sun (that is, emitting 1000 times the power) suddenly goes supernova. Using data from [\[link\]](#): (a) By what factor does its power output increase? (b) How many times brighter than our entire Milky Way galaxy is the supernova? (c) Based on your answers, discuss whether it should be possible to observe supernovas in distant galaxies. Note that there are on the order of 10^{11} observable galaxies, the average brightness of which is somewhat less than our own galaxy.

Exercise:

Problem:

A person in good physical condition can put out 100 W of useful power for several hours at a stretch, perhaps by pedaling a mechanism that drives an electric generator. Neglecting any problems of generator efficiency and practical considerations such as resting time: (a) How many people would it take to run a 4.00-kW electric clothes dryer? (b) How many people would it take to replace a large electric power plant that generates 800 MW?

Solution:

(a) 40

(b) 8 million

Exercise:

Problem:

What is the cost of operating a 3.00-W electric clock for a year if the cost of electricity is \$0.0900 per $\text{kW} \cdot \text{h}$?

Exercise:

Problem:

A large household air conditioner may consume 15.0 kW of power. What is the cost of operating this air conditioner 3.00 h per day for 30.0 d if the cost of electricity is \$0.110 per $\text{kW} \cdot \text{h}$?

Solution:

\$149

Exercise:

Problem:

(a) What is the average power consumption in watts of an appliance that uses 5.00 $\text{kW} \cdot \text{h}$ of energy per day? (b) How many joules of energy does this appliance consume in a year?

Exercise:

Problem:

(a) What is the average useful power output of a person who does $6.00 \times 10^6 \text{ J}$ of useful work in 8.00 h? (b) Working at this rate, how long will it take this person to lift 2000 kg of bricks 1.50 m to a platform? (Work done to lift his body can be omitted because it is not considered useful output here.)

Solution:

(a) 208 W

(b) 141 s

Exercise:

Problem:

A 500-kg dragster accelerates from rest to a final speed of 110 m/s in 400 m (about a quarter of a mile) and encounters an average frictional force of 1200 N. What is its average power output in watts and horsepower if this takes 7.30 s?

Exercise:

Problem:

(a) How long will it take an 850-kg car with a useful power output of 40.0 hp (1 hp = 746 W) to reach a speed of 15.0 m/s, neglecting friction? (b) How long will this acceleration take if the car also climbs a 3.00-m-high hill in the process?

Solution:

(a) 3.20 s

(b) 4.04 s

Exercise:

Problem:

(a) Find the useful power output of an elevator motor that lifts a 2500-kg load a height of 35.0 m in 12.0 s, if it also increases the speed from rest to 4.00 m/s. Note that the total mass of the counterbalanced system is 10,000 kg—so that only 2500 kg is raised in height, but the full 10,000 kg is accelerated. (b) What does it cost, if electricity is \$0.0900 per kW · h?

Exercise:

Problem:

(a) What is the available energy content, in joules, of a battery that operates a 2.00-W electric clock for 18 months? (b) How long can a battery that can supply 8.00×10^4 J run a pocket calculator that consumes energy at the rate of 1.00×10^{-3} W?

Solution:

(a) 9.46×10^7 J

(b) 2.54 y

Exercise:**Problem:**

(a) How long would it take a 1.50×10^5 -kg airplane with engines that produce 100 MW of power to reach a speed of 250 m/s and an altitude of 12.0 km if air resistance were negligible? (b) If it actually takes 900 s, what is the power? (c) Given this power, what is the average force of air resistance if the airplane takes 1200 s? (Hint: You must find the distance the plane travels in 1200 s assuming constant acceleration.)

Exercise:**Problem:**

Calculate the power output needed for a 950-kg car to climb a 2.00° slope at a constant 30.0 m/s while encountering wind resistance and friction totaling 600 N. Explicitly show how you follow the steps in the [Problem-Solving Strategies for Energy](#).

Solution:

Identify knowns: $m = 950$ kg, slope angle $\theta = 2.00^\circ$, $v = 30.0$ m/s, $f = 600$ N

Identify unknowns: power P of the car, force F that car applies to road

Solve for unknown:

$$P = \frac{W}{t} = \frac{Fd}{t} = F\left(\frac{d}{t}\right) = Fv,$$

where F is parallel to the incline and must oppose the resistive forces and the force of gravity:

$$F = f + w = 600 \text{ N} + mg \sin \theta$$

Insert this into the expression for power and solve:

$$\begin{aligned} P &= (f + mg \sin \theta)v \\ &= \left[600 \text{ N} + (950 \text{ kg}) \left(9.80 \text{ m/s}^2 \right) \sin 2^\circ \right] (30.0 \text{ m/s}) \\ &= 2.77 \times 10^4 \text{ W} \end{aligned}$$

About 28 kW (or about 37 hp) is reasonable for a car to climb a gentle incline.

Exercise:

Problem:

(a) Calculate the power per square meter reaching Earth's upper atmosphere from the Sun. (Take the power output of the Sun to be $4.00 \times 10^{26} \text{ W}$.) (b) Part of this is absorbed and reflected by the atmosphere, so that a maximum of 1.30 kW/m^2 reaches Earth's surface. Calculate the area in km^2 of solar energy collectors needed to replace an electric power plant that generates 750 MW if the collectors convert an average of 2.00% of the maximum power into electricity. (This small conversion efficiency is due to the devices themselves, and the fact that the sun is directly overhead only briefly.) With the same assumptions, what area would be needed to meet the United States' energy needs ($1.05 \times 10^{20} \text{ J}$)? Australia's energy needs ($5.4 \times 10^{18} \text{ J}$)? China's energy needs ($6.3 \times 10^{19} \text{ J}$)? (These energy consumption values are from 2006.)

Glossary

power

the rate at which work is done

watt

(W) SI unit of power, with $1 \text{ W} = 1 \text{ J/s}$

horsepower

an older non-SI unit of power, with $1 \text{ hp} = 746 \text{ W}$

kilowatt-hour

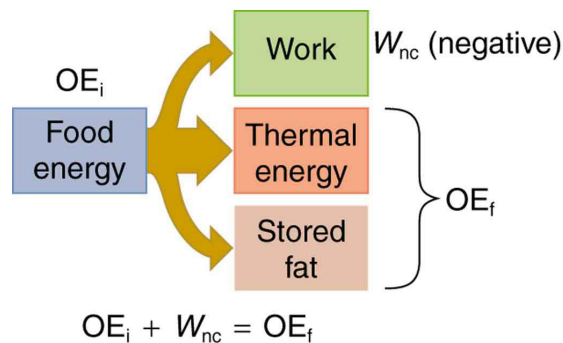
(kW · h) unit used primarily for electrical energy provided by electric utility companies

Work, Energy, and Power in Humans

- Explain the human body's consumption of energy when at rest vs. when engaged in activities that do useful work.
- Calculate the conversion of chemical energy in food into useful work.

Energy Conversion in Humans

Our own bodies, like all living organisms, are energy conversion machines. Conservation of energy implies that the chemical energy stored in food is converted into work, thermal energy, and/or stored as chemical energy in fatty tissue. (See [\[link\]](#).) The fraction going into each form depends both on how much we eat and on our level of physical activity. If we eat more than is needed to do work and stay warm, the remainder goes into body fat.



Energy consumed by humans is converted to work, thermal energy, and stored fat. By far the largest fraction goes to thermal energy, although the fraction varies depending on the type of physical activity.

Power Consumed at Rest

The *rate* at which the body uses food energy to sustain life and to do different activities is called the **metabolic rate**. The total energy conversion rate of a person *at rest* is called the **basal metabolic rate** (BMR) and is divided among various systems in the body, as shown in [\[link\]](#). The largest fraction goes to the liver and spleen, with the brain coming next. Of course, during vigorous exercise, the energy consumption of the skeletal muscles and heart increase markedly. About 75% of the calories burned in a day go into these basic functions. The BMR is a function of age, gender, total body weight, and amount of muscle mass (which burns more calories than body fat). Athletes have a greater BMR due to this last factor.

Organ	Power consumed at rest (W)	Oxygen consumption (mL/min)	Percent of BMR
Liver & spleen	23	67	27
Brain	16	47	19
Skeletal muscle	15	45	18
Kidney	9	26	10
Heart	6	17	7
Other	16	48	19
Totals	85 W	250 mL/min	100%

Basal Metabolic Rates (BMR)

Energy consumption is directly proportional to oxygen consumption because the digestive process is basically one of oxidizing food. We can measure the energy people use during various activities by measuring their oxygen use. (See [\[link\]](#).) Approximately 20 kJ of energy are produced for each liter of oxygen consumed, independent of the type of food. [\[link\]](#) shows energy and oxygen consumption rates (power expended) for a variety of activities.

Power of Doing Useful Work

Work done by a person is sometimes called **useful work**, which is *work done on the outside world*, such as lifting weights. Useful work requires a force exerted through a distance on the outside world, and so it excludes internal work, such as that done by the heart when pumping blood. Useful work does include that done in climbing stairs or accelerating to a full run, because these are accomplished by exerting forces on the outside world. Forces exerted by the body are nonconservative, so that they can change the mechanical energy ($KE + PE$) of the system worked upon, and this is often the goal. A baseball player throwing a ball, for example, increases both the ball's kinetic and potential energy.

If a person needs more energy than they consume, such as when doing vigorous work, the body must draw upon the chemical energy stored in fat. So exercise can be helpful in losing fat. However, the amount of exercise needed to produce a loss in fat, or to burn off extra calories consumed that day, can be large, as [\[link\]](#) illustrates.

Example:

Calculating Weight Loss from Exercising

If a person who normally requires an average of 12,000 kJ (3000 kcal) of food energy per day consumes 13,000 kJ per day, he will steadily gain weight. How much bicycling per day is required to work off this extra 1000 kJ?

Solution

[\[link\]](#) states that 400 W are used when cycling at a moderate speed. The time required to work off 1000 kJ at this rate is then

Equation:

$$\text{Time} = \frac{\text{energy}}{\left(\frac{\text{energy}}{\text{time}}\right)} = \frac{1000 \text{ kJ}}{400 \text{ W}} = 2500 \text{ s} = 42 \text{ min.}$$

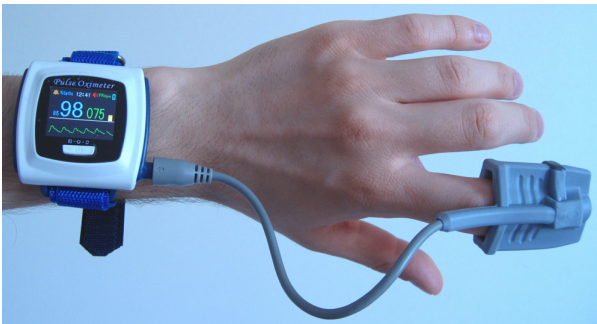
Discussion

If this person uses more energy than he or she consumes, the person's body will obtain the needed energy by metabolizing body fat. If the person uses 13,000 kJ but consumes only 12,000 kJ, then the amount of fat loss will be

Equation:

$$\text{Fat loss} = (1000 \text{ kJ}) \left(\frac{1.0 \text{ g fat}}{39 \text{ kJ}} \right) = 26 \text{ g,}$$

assuming the energy content of fat to be 39 kJ/g.



A pulse oxymeter is an apparatus that measures the amount of oxygen in blood.

Oxymeters can be used to determine a person's metabolic rate, which is the rate at which food energy is converted to another form. Such

measurements can indicate the level of athletic conditioning as well as certain medical problems. (credit: UusiAjaja, Wikimedia Commons)

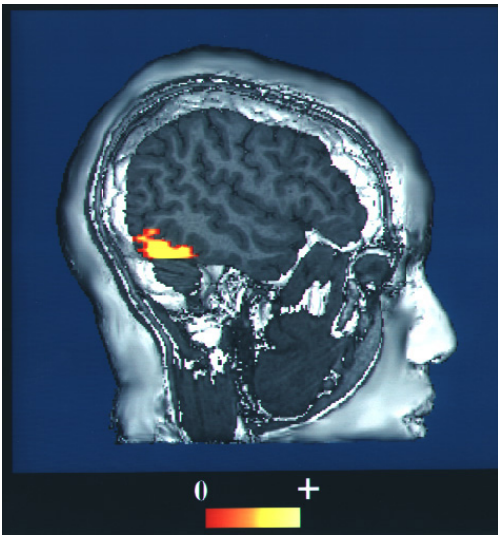
Activity	Energy consumption in watts	Oxygen consumption in liters O₂/min
Sleeping	83	0.24
Sitting at rest	120	0.34
Standing relaxed	125	0.36
Sitting in class	210	0.60
Walking (5 km/h)	280	0.80
Cycling (13–18 km/h)	400	1.14
Shivering	425	1.21
Playing tennis	440	1.26

Activity	Energy consumption in watts	Oxygen consumption in liters O₂/min
Swimming breaststroke	475	1.36
Ice skating (14.5 km/h)	545	1.56
Climbing stairs (116/min)	685	1.96
Cycling (21 km/h)	700	2.00
Running cross-country	740	2.12
Playing basketball	800	2.28
Cycling, professional racer	1855	5.30
Sprinting	2415	6.90

Energy and Oxygen Consumption Rates[\[footnote\]](#) (Power)
for an average 76-kg male

All bodily functions, from thinking to lifting weights, require energy. (See [\[link\]](#).) The many small muscle actions accompanying all quiet activity, from sleeping to head scratching, ultimately become thermal energy, as do less visible muscle actions by the heart, lungs, and digestive tract. Shivering, in fact, is an involuntary response to low body temperature that pits muscles against one another to produce thermal energy in the body (and

do no work). The kidneys and liver consume a surprising amount of energy, but the biggest surprise of all is that a full 25% of all energy consumed by the body is used to maintain electrical potentials in all living cells. (Nerve cells use this electrical potential in nerve impulses.) This bioelectrical energy ultimately becomes mostly thermal energy, but some is utilized to power chemical processes such as in the kidneys and liver, and in fat production.



This fMRI scan shows an increased level of energy consumption in the vision center of the brain. Here, the patient was being asked to recognize faces.
(credit: NIH via Wikimedia Commons)

Section Summary

- The human body converts energy stored in food into work, thermal energy, and/or chemical energy that is stored in fatty tissue.
- The *rate* at which the body uses food energy to sustain life and to do different activities is called the metabolic rate, and the corresponding rate when at rest is called the basal metabolic rate (BMR)
- The energy included in the basal metabolic rate is divided among various systems in the body, with the largest fraction going to the liver and spleen, and the brain coming next.
- About 75% of food calories are used to sustain basic body functions included in the basal metabolic rate.
- The energy consumption of people during various activities can be determined by measuring their oxygen use, because the digestive process is basically one of oxidizing food.

Conceptual Questions

Exercise:

Problem:

Explain why it is easier to climb a mountain on a zigzag path rather than one straight up the side. Is your increase in gravitational potential energy the same in both cases? Is your energy consumption the same in both?

Exercise:

Problem:

Do you do work on the outside world when you rub your hands together to warm them? What is the efficiency of this activity?

Exercise:

Problem:

Shivering is an involuntary response to lowered body temperature. What is the efficiency of the body when shivering, and is this a desirable value?

Exercise:

Problem:

Discuss the relative effectiveness of dieting and exercise in losing weight, noting that most athletic activities consume food energy at a rate of 400 to 500 W, while a single cup of yogurt can contain 1360 kJ (325 kcal). Specifically, is it likely that exercise alone will be sufficient to lose weight? You may wish to consider that regular exercise may increase the metabolic rate, whereas protracted dieting may reduce it.

Problems & Exercises**Exercise:****Problem:**

(a) How long can you rapidly climb stairs (116/min) on the 93.0 kcal of energy in a 10.0-g pat of butter? (b) How many flights is this if each flight has 16 stairs?

Solution:

(a) 9.5 min

(b) 69 flights of stairs

Exercise:**Problem:**

(a) What is the power output in watts and horsepower of a 70.0-kg sprinter who accelerates from rest to 10.0 m/s in 3.00 s? (b) Considering the amount of power generated, do you think a well-trained athlete could do this repetitively for long periods of time?

Exercise:

Problem:

Calculate the power output in watts and horsepower of a shot-putter who takes 1.20 s to accelerate the 7.27-kg shot from rest to 14.0 m/s, while raising it 0.800 m. (Do not include the power produced to accelerate his body.)



Shot putter at the
Dornoch Highland
Gathering in 2007.
(credit: John Haslam,
Flickr)

Solution:

641 W, 0.860 hp

Exercise:**Problem:**

(a) What is the efficiency of an out-of-condition professor who does 2.10×10^5 J of useful work while metabolizing 500 kcal of food energy? (b) How many food calories would a well-conditioned athlete metabolize in doing the same work with an efficiency of 20%?

Exercise:

Problem:

Energy that is not utilized for work or heat transfer is converted to the chemical energy of body fat containing about 39 kJ/g. How many grams of fat will you gain if you eat 10,000 kJ (about 2500 kcal) one day and do nothing but sit relaxed for 16.0 h and sleep for the other 8.00 h? Use data from [\[link\]](#) for the energy consumption rates of these activities.

Solution:

31 g

Exercise:**Problem:**

Using data from [\[link\]](#), calculate the daily energy needs of a person who sleeps for 7.00 h, walks for 2.00 h, attends classes for 4.00 h, cycles for 2.00 h, sits relaxed for 3.00 h, and studies for 6.00 h. (Studying consumes energy at the same rate as sitting in class.)

Exercise:**Problem:**

What is the efficiency of a subject on a treadmill who puts out work at the rate of 100 W while consuming oxygen at the rate of 2.00 L/min? (Hint: See [\[link\]](#).)

Solution:

14.3%

Exercise:

Problem:

Shoveling snow can be extremely taxing because the arms have such a low efficiency in this activity. Suppose a person shoveling a footpath metabolizes food at the rate of 800 W. (a) What is her useful power output? (b) How long will it take her to lift 3000 kg of snow 1.20 m? (This could be the amount of heavy snow on 20 m of footpath.) (c) How much waste heat transfer in kilojoules will she generate in the process?

Exercise:**Problem:**

Very large forces are produced in joints when a person jumps from some height to the ground. (a) Calculate the magnitude of the force produced if an 80.0-kg person jumps from a 0.600-m-high ledge and lands stiffly, compressing joint material 1.50 cm as a result. (Be certain to include the weight of the person.) (b) In practice the knees bend almost involuntarily to help extend the distance over which you stop. Calculate the magnitude of the force produced if the stopping distance is 0.300 m. (c) Compare both forces with the weight of the person.

Solution:

(a) $3.21 \times 10^4 \text{ N}$

(b) $2.35 \times 10^3 \text{ N}$

(c) Ratio of net force to weight of person is 41.0 in part (a); 3.00 in part (b)

Exercise:

Problem:

Jogging on hard surfaces with insufficiently padded shoes produces large forces in the feet and legs. (a) Calculate the magnitude of the force needed to stop the downward motion of a jogger's leg, if his leg has a mass of 13.0 kg, a speed of 6.00 m/s, and stops in a distance of 1.50 cm. (Be certain to include the weight of the 75.0-kg jogger's body.) (b) Compare this force with the weight of the jogger.

Exercise:**Problem:**

(a) Calculate the energy in kJ used by a 55.0-kg woman who does 50 deep knee bends in which her center of mass is lowered and raised 0.400 m. (She does work in both directions.) You may assume her efficiency is 20%. (b) What is the average power consumption rate in watts if she does this in 3.00 min?

Solution:

(a) 108 kJ

(b) 599 W

Exercise:**Problem:**

Kanellos Kanellopoulos flew 119 km from Crete to Santorini, Greece, on April 23, 1988, in the *Daedalus 88*, an aircraft powered by a bicycle-type drive mechanism (see [\[link\]](#)). His useful power output for the 234-min trip was about 350 W. Using the efficiency for cycling from [\[link\]](#), calculate the food energy in kilojoules he metabolized during the flight.

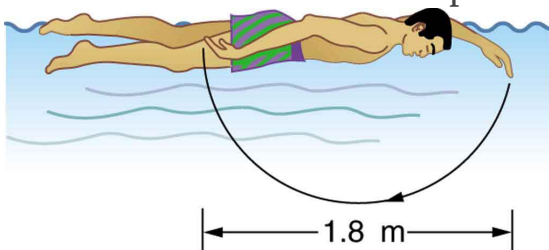


The Daedalus 88 in flight.
(credit: NASA photo by
Beasley)

Exercise:

Problem:

The swimmer shown in [\[link\]](#) exerts an average horizontal backward force of 80.0 N with his arm during each 1.80 m long stroke. (a) What is his work output in each stroke? (b) Calculate the power output of his arms if he does 120 strokes per minute.



Solution:

(a) 144 J

(b) 288 W

Exercise:

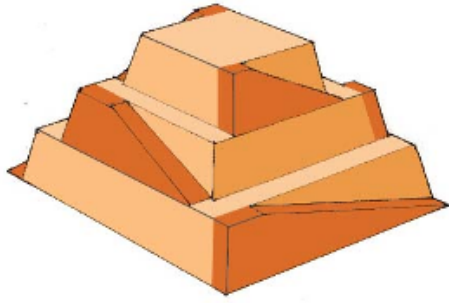
Problem:

Mountain climbers carry bottled oxygen when at very high altitudes.

(a) Assuming that a mountain climber uses oxygen at twice the rate for climbing 116 stairs per minute (because of low air temperature and winds), calculate how many liters of oxygen a climber would need for 10.0 h of climbing. (These are liters at sea level.) Note that only 40% of the inhaled oxygen is utilized; the rest is exhaled. (b) How much useful work does the climber do if he and his equipment have a mass of 90.0 kg and he gains 1000 m of altitude? (c) What is his efficiency for the 10.0-h climb?

Exercise:**Problem:**

The awe-inspiring Great Pyramid of Cheops was built more than 4500 years ago. Its square base, originally 230 m on a side, covered 13.1 acres, and it was 146 m high, with a mass of about 7×10^9 kg. (The pyramid's dimensions are slightly different today due to quarrying and some sagging.) Historians estimate that 20,000 workers spent 20 years to construct it, working 12-hour days, 330 days per year. (a) Calculate the gravitational potential energy stored in the pyramid, given its center of mass is at one-fourth its height. (b) Only a fraction of the workers lifted blocks; most were involved in support services such as building ramps (see [\[link\]](#)), bringing food and water, and hauling blocks to the site. Calculate the efficiency of the workers who did the lifting, assuming there were 1000 of them and they consumed food energy at the rate of 300 kcal/h. What does your answer imply about how much of their work went into block-lifting, versus how much work went into friction and lifting and lowering their own bodies? (c) Calculate the mass of food that had to be supplied each day, assuming that the average worker required 3600 kcal per day and that their diet was 5% protein, 60% carbohydrate, and 35% fat. (These proportions neglect the mass of bulk and nondigestible materials consumed.)



Ancient pyramids were probably constructed using ramps as simple machines.
(credit: Franck Monnier, Wikimedia Commons)

Solution:

- (a) $2.50 \times 10^{12} \text{ J}$
- (b) 2.52%
- (c) $1.4 \times 10^4 \text{ kg}$ (14 metric tons)

Exercise:

Problem:

(a) How long can you play tennis on the 800 kJ (about 200 kcal) of energy in a candy bar? (b) Does this seem like a long time? Discuss why exercise is necessary but may not be sufficient to cause a person to lose weight.

Glossary

metabolic rate

the rate at which the body uses food energy to sustain life and to do different activities

basal metabolic rate

the total energy conversion rate of a person at rest

useful work

work done on an external system

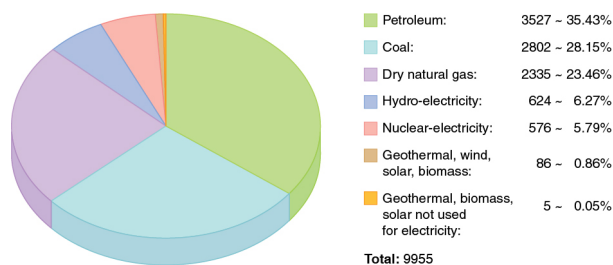
World Energy Use

- Describe the distinction between renewable and nonrenewable energy sources.
- Explain why the inevitable conversion of energy to less useful forms makes it necessary to conserve energy resources.

Energy is an important ingredient in all phases of society. We live in a very interdependent world, and access to adequate and reliable energy resources is crucial for economic growth and for maintaining the quality of our lives. But current levels of energy consumption and production are not sustainable. About 40% of the world's energy comes from oil, and much of that goes to transportation uses. Oil prices are dependent as much upon new (or foreseen) discoveries as they are upon political events and situations around the world. The U.S., with 4.5% of the world's population, consumes 24% of the world's oil production per year; 66% of that oil is imported!

Renewable and Nonrenewable Energy Sources

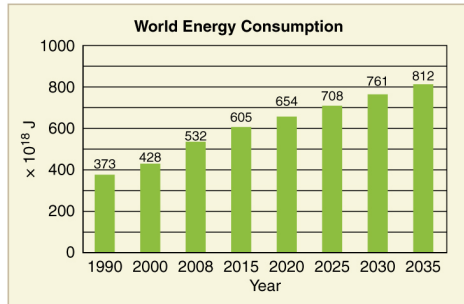
The principal energy resources used in the world are shown in [\[link\]](#). The fuel mix has changed over the years but now is dominated by oil, although natural gas and solar contributions are increasing. **Renewable forms of energy** are those sources that cannot be used up, such as water, wind, solar, and biomass. About 85% of our energy comes from nonrenewable **fossil fuels**—oil, natural gas, coal. The likelihood of a link between global warming and fossil fuel use, with its production of carbon dioxide through combustion, has made, in the eyes of many scientists, a shift to non-fossil fuels of utmost importance—but it will not be easy.



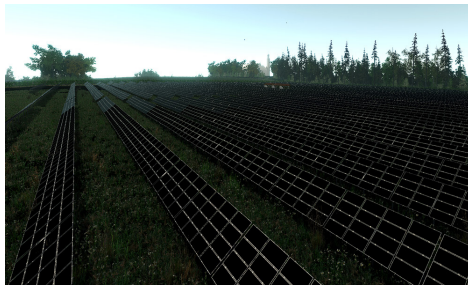
World energy consumption by source, in billions of kilowatt-hours: 2006. (credit: KVDP)

The World's Growing Energy Needs

World energy consumption continues to rise, especially in the developing countries. (See [\[link\]](#).) Global demand for energy has tripled in the past 50 years and might triple again in the next 30 years. While much of this growth will come from the rapidly booming economies of China and India, many of the developed countries, especially those in Europe, are hoping to meet their energy needs by expanding the use of renewable sources. Although presently only a small percentage, renewable energy is growing very fast, especially wind energy. For example, Germany plans to meet 20% of its electricity and 10% of its overall energy needs with renewable resources by the year 2020. (See [\[link\]](#).) Energy is a key constraint in the rapid economic growth of China and India. In 2003, China surpassed Japan as the world's second largest consumer of oil. However, over 1/3 of this is imported. Unlike most Western countries, coal dominates the commercial energy resources of China, accounting for 2/3 of its energy consumption. In 2009 China surpassed the United States as the largest generator of CO₂. In India, the main energy resources are biomass (wood and dung) and coal. Half of India's oil is imported. About 70% of India's electricity is generated by highly polluting coal. Yet there are sizeable strides being made in renewable energy. India has a rapidly growing wind energy base, and it has the largest solar cooking program in the world.



Past and projected world energy use
(source: Based on data from U.S.
Energy Information Administration,
2011)



Solar cell arrays at a power plant in
Steindorf, Germany (credit: Michael
Betke, Flickr)

[\[link\]](#) displays the 2006 commercial energy mix by country for some of the prime energy users in the world. While non-renewable sources dominate, some countries get a sizeable percentage of their electricity from renewable resources. For example, about 67% of New Zealand's electricity demand is met by hydroelectric. Only 10% of the U.S. electricity is generated by renewable resources, primarily hydroelectric. It is difficult to determine total contributions of renewable energy in some countries with a large rural population, so these percentages in this table are left blank.

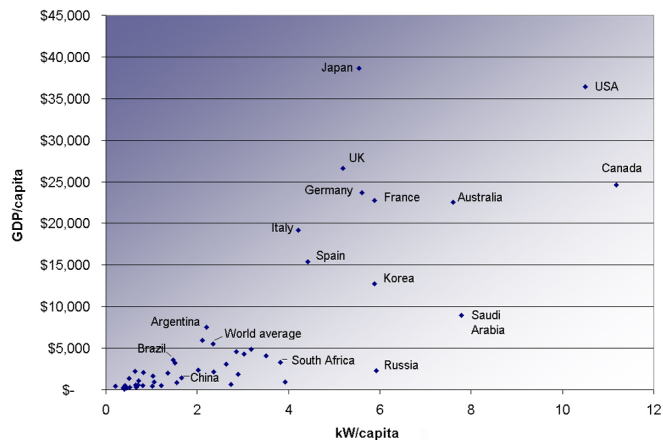
Country	Consumption, in EJ (10 ¹⁸ J)	Oil	Natural Gas	Coal	Nuclear	Hydro	Other Renewables
Australia	5.4	34%	17%	44%	0%	3%	1%

Country	Consumption, in EJ (10^{18} J)	Oil	Natural Gas	Coal	Nuclear	Hydro	Other Renewables
Brazil	9.6	48%	7%	5%	1%	35%	2%
China	63	22%	3%	69%	1%	6%	
Egypt	2.4	50%	41%	1%	0%	6%	
Germany	16	37%	24%	24%	11%	1%	3%
India	15	34%	7%	52%	1%	5%	
Indonesia	4.9	51%	26%	16%	0%	2%	3%
Japan	24	48%	14%	21%	12%	4%	1%
New Zealand	0.44	32%	26%	6%	0%	11%	19%
Russia	31	19%	53%	16%	5%	6%	
U.S.	105	40%	23%	22%	8%	3%	1%
World	432	39%	23%	24%	6%	6%	2%

Energy Consumption—Selected Countries (2006)

Energy and Economic Well-being

The last two columns in this table examine the energy and electricity use per capita. Economic well-being is dependent upon energy use, and in most countries higher standards of living, as measured by GDP (gross domestic product) per capita, are matched by higher levels of energy consumption per capita. This is borne out in [\[link\]](#). Increased efficiency of energy use will change this dependency. A global problem is balancing energy resource development against the harmful effects upon the environment in its extraction and use.



Power consumption per capita versus GDP per capita for various countries. Note the increase in energy usage with increasing GDP. (2007, credit: Frank van Mierlo, Wikimedia Commons)

Conserving Energy

As we finish this chapter on energy and work, it is relevant to draw some distinctions between two sometimes misunderstood terms in the area of energy use. As has been mentioned elsewhere, the “law of the conservation of energy” is a very useful principle in analyzing physical processes. It is a statement that cannot be proven from basic principles, but is a very good bookkeeping device, and no exceptions have ever been found. It states that the total amount of energy in an isolated system will always remain constant. Related to this principle, but remarkably different from it, is the important philosophy of energy conservation. This concept has to do with seeking to decrease the amount of energy used by an individual or group through (1) reduced activities (e.g., turning down thermostats, driving fewer kilometers) and/or (2) increasing conversion efficiencies in the performance of a particular task—such as developing and using more efficient room heaters, cars that have greater miles-per-gallon ratings, energy-efficient compact fluorescent lights, etc.

Since energy in an isolated system is not destroyed or created or generated, one might wonder why we need to be concerned about our energy resources, since energy is a conserved quantity. The problem is that the final result of most energy transformations is waste heat transfer to the environment and conversion to energy forms no longer useful for doing work. To state it in another way, the potential for energy to produce useful work has been “degraded” in the energy transformation. (This will be discussed in more detail in [Thermodynamics](#).)

Section Summary

- The relative use of different fuels to provide energy has changed over the years, but fuel use is currently dominated by oil, although natural gas and solar contributions are increasing.
- Although non-renewable sources dominate, some countries meet a sizeable percentage of their electricity needs from renewable resources.
- The United States obtains only about 10% of its energy from renewable sources, mostly hydroelectric power.
- Economic well-being is dependent upon energy use, and in most countries higher standards of living, as measured by GDP (Gross Domestic Product) per capita, are matched by higher levels of energy consumption per capita.
- Even though, in accordance with the law of conservation of energy, energy can never be created or destroyed, energy that can be used to do work is always partly converted to less useful forms, such as waste heat to the environment, in all of our uses of energy for practical purposes.

Conceptual Questions

Exercise:

Problem:

What is the difference between energy conservation and the law of conservation of energy? Give some examples of each.

Exercise:

Problem:

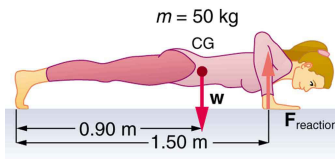
If the efficiency of a coal-fired electrical generating plant is 35%, then what do we mean when we say that energy is a conserved quantity?

Problems & Exercises

Exercise:

Problem: Integrated Concepts

(a) Calculate the force the woman in [\[link\]](#) exerts to do a push-up at constant speed, taking all data to be known to three digits. (b) How much work does she do if her center of mass rises 0.240 m? (c) What is her useful power output if she does 25 push-ups in 1 min? (Should work done lowering her body be included? See the discussion of useful work in [Work, Energy, and Power in Humans](#).



Forces involved in doing push-ups. The woman's weight acts as a force exerted downward on her center of gravity (CG).

Solution:

- (a) 294 N
- (b) 118 J
- (c) 49.0 W

Exercise:

Problem: Integrated Concepts

A 75.0-kg cross-country skier is climbing a 3.0° slope at a constant speed of 2.00 m/s and encounters air resistance of 25.0 N. Find his power output for work done against the gravitational force and air resistance. (b) What average force does he exert backward on the snow to accomplish this? (c) If he continues to exert this force and to experience the same air resistance when he reaches a level area, how long will it take him to reach a velocity of 10.0 m/s?

Exercise:**Problem: Integrated Concepts**

The 70.0-kg swimmer in [\[link\]](#) starts a race with an initial velocity of 1.25 m/s and exerts an average force of 80.0 N backward with his arms during each 1.80 m long stroke. (a) What is his initial acceleration if water resistance is 45.0 N? (b) What is the subsequent average resistance force from the water during the 5.00 s it takes him to reach his top velocity of 2.50 m/s? (c) Discuss whether water resistance seems to increase linearly with velocity.

Solution:

(a) 0.500 m/s^2

(b) 62.5 N

(c) Assuming the acceleration of the swimmer decreases linearly with time over the 5.00 s interval, the frictional force must therefore be increasing linearly with time, since $f = F - ma$. If the acceleration decreases linearly with time, the velocity will contain a term dependent on time squared (t^2). Therefore, the water resistance will not depend linearly on the velocity.

Exercise:**Problem: Integrated Concepts**

A toy gun uses a spring with a force constant of 300 N/m to propel a 10.0-g steel ball. If the spring is compressed 7.00 cm and friction is negligible: (a) How much force is needed to compress the spring? (b) To what maximum height can the ball be shot? (c) At what angles above the horizontal may a child aim to hit a target 3.00 m away at the same height as the gun? (d) What is the gun's maximum range on level ground?

Exercise:**Problem: Integrated Concepts**

(a) What force must be supplied by an elevator cable to produce an acceleration of 0.800 m/s^2 against a 200-N frictional force, if the mass of the loaded elevator is 1500 kg? (b) How much work is done by the cable in lifting the elevator 20.0 m? (c) What is the final speed of the elevator if it starts from rest? (d) How much work went into thermal energy?

Solution:

(a) $16.1 \times 10^3 \text{ N}$

(b) $3.22 \times 10^5 \text{ J}$

(c) 5.66 m/s

(d) 4.00 kJ

Exercise:**Problem: Unreasonable Results**

A car advertisement claims that its 900-kg car accelerated from rest to 30.0 m/s and drove 100 km, gaining 3.00 km in altitude, on 1.0 gal of gasoline. The average force of friction including air resistance was 700 N. Assume all values are known to three significant figures. (a) Calculate the car's efficiency. (b) What is unreasonable about the result? (c) Which premise is unreasonable, or which premises are inconsistent?

Exercise:**Problem: Unreasonable Results**

Body fat is metabolized, supplying 9.30 kcal/g, when dietary intake is less than needed to fuel metabolism. The manufacturers of an exercise bicycle claim that you can lose 0.500 kg of fat per day by vigorously exercising for 2.00 h per day on their machine. (a) How many kcal are supplied by the metabolization of 0.500 kg of fat? (b) Calculate the kcal/min that you would have to utilize to metabolize fat at the rate of 0.500 kg in 2.00 h. (c) What is unreasonable about the results? (d) Which premise is unreasonable, or which premises are inconsistent?

Solution:

(a) 4.65×10^3 kcal

(b) 38.8 kcal/min

(c) This power output is higher than the highest value on [\[link\]](#), which is about 35 kcal/min (corresponding to 2415 watts) for sprinting.

(d) It would be impossible to maintain this power output for 2 hours (imagine sprinting for 2 hours!).

Exercise:**Problem: Construct Your Own Problem**

Consider a person climbing and descending stairs. Construct a problem in which you calculate the long-term rate at which stairs can be climbed considering the mass of the person, his ability to generate power with his legs, and the height of a single stair step. Also consider why the same person can descend stairs at a faster rate for a nearly unlimited time in spite of the fact that very similar forces are exerted going down as going up. (This points to a fundamentally different process for descending versus climbing stairs.)

Exercise:**Problem: Construct Your Own Problem**

Consider humans generating electricity by pedaling a device similar to a stationary bicycle. Construct a problem in which you determine the number of people it would take to replace a large electrical generation facility. Among the things to consider are the power output that is reasonable using the legs, rest time, and the need for electricity 24 hours per day. Discuss the practical implications of your results.

Exercise:**Problem: Integrated Concepts**

A 105-kg basketball player crouches down 0.400 m while waiting to jump. After exerting a force on the floor through this 0.400 m, his feet leave the floor and his center of gravity rises 0.950 m above its normal standing erect position. (a) Using energy considerations, calculate his velocity when he leaves the floor. (b) What average force did he exert on the floor? (Do not neglect the force to support his weight as well as that to accelerate him.) (c) What was his power output during the acceleration phase?

Solution:

(a) 4.32 m/s

(b) 3.47×10^3 N

(c) 8.93 kW

Glossary

renewable forms of energy

those sources that cannot be used up, such as water, wind, solar, and biomass

fossil fuels

oil, natural gas, and coal

Introduction to Linear Momentum and Collisions

class="introduction"

"Each
rugby
player has
great
momentum
, which will
affect the
outcome of
their
collisions
with each
other and
the ground.
(credit:
vjpaul,
Flickr)"



We use the term momentum in various ways in everyday language, and most of these ways are consistent with its precise scientific definition. We speak of sports teams or politicians gaining and maintaining the momentum to win. We also recognize that momentum has something to do with collisions. For example, looking at the rugby players in the photograph colliding and falling to the ground, we expect their momenta to have great effects in the resulting collisions. Generally, momentum implies a tendency to continue on course—to move in the same direction—and is associated with great mass and speed.

Momentum, like energy, is important because it is conserved. Only a few physical quantities are conserved in nature, and studying them yields fundamental insight into how nature works, as we shall see in our study of momentum.

Linear Momentum and Force

- Define linear momentum.
- Explain the relationship between momentum and force.
- State Newton's second law of motion in terms of momentum.
- Calculate momentum given mass and velocity.

Linear Momentum

The scientific definition of linear momentum is consistent with most people's intuitive understanding of momentum: a large, fast-moving object has greater momentum than a smaller, slower object. **Linear momentum** is defined as the product of a system's mass multiplied by its velocity. In symbols, linear momentum is expressed as

Equation:

$$\mathbf{p} = m\mathbf{v}.$$

Momentum is directly proportional to the object's mass and also its velocity. Thus the greater an object's mass or the greater its velocity, the greater its momentum. Momentum \mathbf{p} is a vector having the same direction as the velocity \mathbf{v} . The SI unit for momentum is $\text{kg} \cdot \text{m/s}$.

Note:

Linear Momentum

Linear momentum is defined as the product of a system's mass multiplied by its velocity:

Equation:

$$\mathbf{p} = m\mathbf{v}.$$

Example:

Calculating Momentum: A Football Player and a Football

(a) Calculate the momentum of a 110-kg football player running at 8.00 m/s. (b) Compare the player's momentum with the momentum of a hard-thrown 0.410-kg football that has a speed of 25.0 m/s.

Strategy

No information is given regarding direction, and so we can calculate only the magnitude of the momentum, p . (As usual, a symbol that is in italics is a magnitude, whereas one that is italicized, boldfaced, and has an arrow is a vector.) In both parts of this example, the magnitude of momentum can be calculated directly from the definition of momentum given in the equation, which becomes

Equation:

$$p = mv$$

when only magnitudes are considered.

Solution for (a)

To determine the momentum of the player, substitute the known values for the player's mass and speed into the equation.

Equation:

$$p_{\text{player}} = (110 \text{ kg})(8.00 \text{ m/s}) = 880 \text{ kg} \cdot \text{m/s}$$

Solution for (b)

To determine the momentum of the ball, substitute the known values for the ball's mass and speed into the equation.

Equation:

$$p_{\text{ball}} = (0.410 \text{ kg})(25.0 \text{ m/s}) = 10.3 \text{ kg} \cdot \text{m/s}$$

The ratio of the player's momentum to that of the ball is

Equation:

$$\frac{p_{\text{player}}}{p_{\text{ball}}} = \frac{880}{10.3} = 85.9.$$

Discussion

Although the ball has greater velocity, the player has a much greater mass. Thus the momentum of the player is much greater than the momentum of the football, as you might guess. As a result, the player's motion is only slightly affected if he catches the ball. We shall quantify what happens in such collisions in terms of momentum in later sections.

Momentum and Newton's Second Law

The importance of momentum, unlike the importance of energy, was recognized early in the development of classical physics. Momentum was deemed so important that it was called the “quantity of motion.” Newton actually stated his **second law of motion** in terms of momentum: The net external force equals the change in momentum of a system divided by the time over which it changes. Using symbols, this law is

Equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t},$$

where \mathbf{F}_{net} is the net external force, $\Delta \mathbf{p}$ is the change in momentum, and Δt is the change in time.

Note:

Newton's Second Law of Motion in Terms of Momentum

The net external force equals the change in momentum of a system divided by the time over which it changes.

Equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$$

Note:**Making Connections: Force and Momentum**

Force and momentum are intimately related. Force acting over time can change momentum, and Newton's second law of motion, can be stated in its most broadly applicable form in terms of momentum. Momentum continues to be a key concept in the study of atomic and subatomic particles in quantum mechanics.

This statement of Newton's second law of motion includes the more familiar $\mathbf{F}_{\text{net}} = m\mathbf{a}$ as a special case. We can derive this form as follows. First, note that the change in momentum $\Delta\mathbf{p}$ is given by

Equation:

$$\Delta\mathbf{p} = \Delta(m\mathbf{v}).$$

If the mass of the system is constant, then

Equation:

$$\Delta(m\mathbf{v}) = m\Delta\mathbf{v}.$$

So that for constant mass, Newton's second law of motion becomes

Equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta\mathbf{p}}{\Delta t} = \frac{m\Delta\mathbf{v}}{\Delta t}.$$

Because $\frac{\Delta\mathbf{v}}{\Delta t} = \mathbf{a}$, we get the familiar equation

Equation:

$$\mathbf{F}_{\text{net}} = m\mathbf{a}$$

when the mass of the system is constant.

Newton's second law of motion stated in terms of momentum is more generally applicable because it can be applied to systems where the mass is changing, such as rockets, as well as to systems of constant mass. We will consider systems with varying mass in some detail; however, the relationship between momentum and force remains useful when mass is constant, such as in the following example.

Example:

Calculating Force: Venus Williams' Racquet

During the 2007 French Open, Venus Williams hit the fastest recorded serve in a premier women's match, reaching a speed of 58 m/s (209 km/h). What is the average force exerted on the 0.057-kg tennis ball by Venus Williams' racquet, assuming that the ball's speed just after impact is 58 m/s, that the initial horizontal component of the velocity before impact is negligible, and that the ball remained in contact with the racquet for 5.0 ms (milliseconds)?

Strategy

This problem involves only one dimension because the ball starts from having no horizontal velocity component before impact. Newton's second law stated in terms of momentum is then written as

Equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}.$$

As noted above, when mass is constant, the change in momentum is given by

Equation:

$$\Delta p = m\Delta v = m(v_f - v_i).$$

In this example, the velocity just after impact and the change in time are given; thus, once Δp is calculated, $F_{\text{net}} = \frac{\Delta p}{\Delta t}$ can be used to find the force.

Solution

To determine the change in momentum, substitute the values for the initial and final velocities into the equation above.

Equation:

$$\begin{aligned}\Delta p &= m(v_f - v_i) \\ &= (0.057 \text{ kg})(58 \text{ m/s} - 0 \text{ m/s}) \\ &= 3.306 \text{ kg} \cdot \text{m/s} \approx 3.3 \text{ kg} \cdot \text{m/s}\end{aligned}$$

Now the magnitude of the net external force can be determined by using

$$F_{\text{net}} = \frac{\Delta p}{\Delta t}:$$

Equation:

$$\begin{aligned}F_{\text{net}} &= \frac{\Delta p}{\Delta t} = \frac{3.306 \text{ kg} \cdot \text{m/s}}{5.0 \times 10^{-3} \text{ s}} \\ &= 661 \text{ N} \approx 660 \text{ N},\end{aligned}$$

where we have retained only two significant figures in the final step.

Discussion

This quantity was the average force exerted by Venus Williams' racquet on the tennis ball during its brief impact (note that the ball also experienced the 0.56-N force of gravity, but that force was not due to the racquet). This problem could also be solved by first finding the acceleration and then using $F_{\text{net}} = ma$, but one additional step would be required compared with the strategy used in this example.

Section Summary

- Linear momentum (*momentum* for brevity) is defined as the product of a system's mass multiplied by its velocity.
- In symbols, linear momentum **p** is defined to be

Equation:

$$\mathbf{p} = m\mathbf{v},$$

where m is the mass of the system and \mathbf{v} is its velocity.

- The SI unit for momentum is $\text{kg} \cdot \text{m/s}$.

- Newton's second law of motion in terms of momentum states that the net external force equals the change in momentum of a system divided by the time over which it changes.
- In symbols, Newton's second law of motion is defined to be

Equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t},$$

\mathbf{F}_{net} is the net external force, $\Delta \mathbf{p}$ is the change in momentum, and Δt is the change time.

Conceptual Questions

Exercise:

Problem:

An object that has a small mass and an object that has a large mass have the same momentum. Which object has the largest kinetic energy?

Exercise:

Problem:

An object that has a small mass and an object that has a large mass have the same kinetic energy. Which mass has the largest momentum?

Exercise:

Problem: Professional Application

Football coaches advise players to block, hit, and tackle with their feet on the ground rather than by leaping through the air. Using the concepts of momentum, work, and energy, explain how a football player can be more effective with his feet on the ground.

Exercise:

Problem:

How can a small force impart the same momentum to an object as a large force?

Problems & Exercises**Exercise:****Problem:**

(a) Calculate the momentum of a 2000-kg elephant charging a hunter at a speed of 7.50 m/s. (b) Compare the elephant's momentum with the momentum of a 0.0400-kg tranquilizer dart fired at a speed of 600 m/s. (c) What is the momentum of the 90.0-kg hunter running at 7.40 m/s after missing the elephant?

Solution:

(a) $1.50 \times 10^4 \text{ kg} \cdot \text{m/s}$

(b) 625 to 1

(c) $6.66 \times 10^2 \text{ kg} \cdot \text{m/s}$

Exercise:**Problem:**

(a) What is the mass of a large ship that has a momentum of $1.60 \times 10^9 \text{ kg} \cdot \text{m/s}$, when the ship is moving at a speed of 48.0 km/h? (b) Compare the ship's momentum to the momentum of a 1100-kg artillery shell fired at a speed of 1200 m/s.

Exercise:

Problem:

(a) At what speed would a 2.00×10^4 -kg airplane have to fly to have a momentum of $1.60 \times 10^9 \text{ kg} \cdot \text{m/s}$ (the same as the ship's momentum in the problem above)? (b) What is the plane's momentum when it is taking off at a speed of 60.0 m/s ? (c) If the ship is an aircraft carrier that launches these airplanes with a catapult, discuss the implications of your answer to (b) as it relates to recoil effects of the catapult on the ship.

Solution:

(a) $8.00 \times 10^4 \text{ m/s}$

(b) $1.20 \times 10^6 \text{ kg} \cdot \text{m/s}$

(c) Because the momentum of the airplane is 3 orders of magnitude smaller than of the ship, the ship will not recoil very much. The recoil would be -0.0100 m/s , which is probably not noticeable.

Exercise:**Problem:**

(a) What is the momentum of a garbage truck that is $1.20 \times 10^4 \text{ kg}$ and is moving at 10.0 m/s ? (b) At what speed would an 8.00 -kg trash can have the same momentum as the truck?

Exercise:**Problem:**

A runaway train car that has a mass of $15,000 \text{ kg}$ travels at a speed of 5.4 m/s down a track. Compute the time required for a force of 1500 N to bring the car to rest.

Solution:

54 s

Exercise:**Problem:**

The mass of Earth is 5.972×10^{24} kg and its orbital radius is an average of 1.496×10^{11} m. Calculate its linear momentum.

Glossary

linear momentum

the product of mass and velocity

second law of motion

physical law that states that the net external force equals the change in momentum of a system divided by the time over which it changes

Impulse

- Define impulse.
- Describe effects of impulses in everyday life.
- Determine the average effective force using graphical representation.
- Calculate average force and impulse given mass, velocity, and time.

The effect of a force on an object depends on how long it acts, as well as how great the force is. In [\[link\]](#), a very large force acting for a short time had a great effect on the momentum of the tennis ball. A small force could cause the same **change in momentum**, but it would have to act for a much longer time. For example, if the ball were thrown upward, the gravitational force (which is much smaller than the tennis racquet's force) would eventually reverse the momentum of the ball. Quantitatively, the effect we are talking about is the change in momentum $\Delta \mathbf{p}$.

By rearranging the equation $\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$ to be

Equation:

$$\Delta \mathbf{p} = \mathbf{F}_{\text{net}} \Delta t,$$

we can see how the change in momentum equals the average net external force multiplied by the time this force acts. The quantity $\mathbf{F}_{\text{net}} \Delta t$ is given the name **impulse**. Impulse is the same as the change in momentum.

Note:

Impulse: Change in Momentum

Change in momentum equals the average net external force multiplied by the time this force acts.

Equation:

$$\Delta \mathbf{p} = \mathbf{F}_{\text{net}} \Delta t$$

The quantity $\mathbf{F}_{\text{net}} \Delta t$ is given the name impulse.

There are many ways in which an understanding of impulse can save lives, or at least limbs. The dashboard padding in a car, and certainly the airbags, allow the net force on the occupants in the car to act over a much longer time when there is a sudden stop. The momentum change is the same for an occupant, whether an air bag is deployed or not, but the force (to bring the occupant to a stop) will be much less if it acts over a larger time. Cars today have many plastic components. One advantage of plastics is their lighter weight, which results in better gas mileage. Another advantage is that a car will crumple in a collision, especially in the event of a head-on collision. A longer collision time means the force on the car will be less. Deaths during car races decreased dramatically when the rigid frames of racing cars were replaced with parts that could crumple or collapse in the event of an accident.

Bones in a body will fracture if the force on them is too large. If you jump onto the floor from a table, the force on your legs can be immense if you land stiff-legged on a hard surface. Rolling on the ground after jumping from the table, or landing with a parachute, extends the time over which the force (on you from the ground) acts.

Example:

Calculating Magnitudes of Impulses: Two Billiard Balls Striking a Rigid Wall

Two identical billiard balls strike a rigid wall with the same speed, and are reflected without any change of speed. The first ball strikes perpendicular to the wall. The second ball strikes the wall at an angle of 30° from the perpendicular, and bounces off at an angle of 30° from perpendicular to the wall.

- (a) Determine the direction of the force on the wall due to each ball.
- (b) Calculate the ratio of the magnitudes of impulses on the two balls by the wall.

Strategy for (a)

In order to determine the force on the wall, consider the force on the ball due to the wall using Newton's second law and then apply Newton's third law to determine the direction. Assume the x -axis to be normal to the wall and to be positive in the initial direction of motion. Choose the y -axis to be

along the wall in the plane of the second ball's motion. The momentum direction and the velocity direction are the same.

Solution for (a)

The first ball bounces directly into the wall and exerts a force on it in the $+x$ direction. Therefore the wall exerts a force on the ball in the $-x$ direction. The second ball continues with the same momentum component in the y direction, but reverses its x -component of momentum, as seen by sketching a diagram of the angles involved and keeping in mind the proportionality between velocity and momentum.

These changes mean the change in momentum for both balls is in the $-x$ direction, so the force of the wall on each ball is along the $-x$ direction.

Strategy for (b)

Calculate the change in momentum for each ball, which is equal to the impulse imparted to the ball.

Solution for (b)

Let u be the speed of each ball before and after collision with the wall, and m the mass of each ball. Choose the x -axis and y -axis as previously described, and consider the change in momentum of the first ball which strikes perpendicular to the wall.

Equation:

$$p_{xi} = mu; p_{yi} = 0$$

Equation:

$$p_{xf} = -mu; p_{yf} = 0$$

Impulse is the change in momentum vector. Therefore the x -component of impulse is equal to $-2mu$ and the y -component of impulse is equal to zero.

Now consider the change in momentum of the second ball.

Equation:

$$p_{xi} = mu \cos 30^\circ; p_{yi} = -mu \sin 30^\circ$$

Equation:

$$p_{xf} = -mu \cos 30^\circ; p_{yf} = -mu \sin 30^\circ$$

It should be noted here that while p_x changes sign after the collision, p_y does not. Therefore the x -component of impulse is equal to $-2mu \cos 30^\circ$ and the y -component of impulse is equal to zero.

The ratio of the magnitudes of the impulse imparted to the balls is

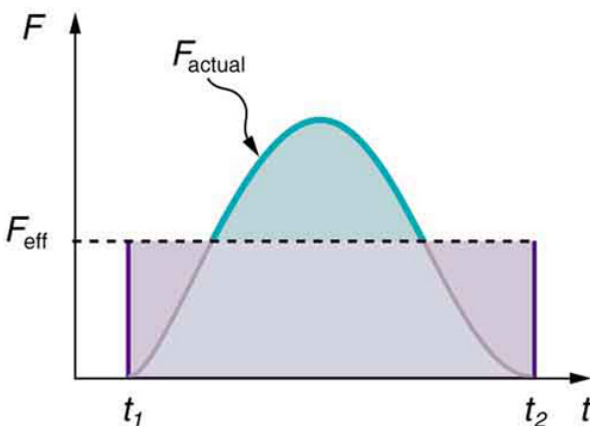
Equation:

$$\frac{2mu}{2mu \cos 30^\circ} = \frac{2}{\sqrt{3}} = 1.155.$$

Discussion

The direction of impulse and force is the same as in the case of (a); it is normal to the wall and along the negative x -direction. Making use of Newton's third law, the force on the wall due to each ball is normal to the wall along the positive x -direction.

Our definition of impulse includes an assumption that the force is constant over the time interval Δt . *Forces are usually not constant.* Forces vary considerably even during the brief time intervals considered. It is, however, possible to find an average effective force F_{eff} that produces the same result as the corresponding time-varying force. [\[link\]](#) shows a graph of what an actual force looks like as a function of time for a ball bouncing off the floor. The area under the curve has units of momentum and is equal to the impulse or change in momentum between times t_1 and t_2 . That area is equal to the area inside the rectangle bounded by F_{eff} , t_1 , and t_2 . Thus the impulses and their effects are the same for both the actual and effective forces.



A graph of force versus time with time along the x -axis and force along the y -axis for an actual force and an equivalent effective force. The areas under the two curves are equal.

Note:

Making Connections: Take-Home Investigation—Hand Movement and Impulse

Try catching a ball while “giving” with the ball, pulling your hands toward your body. Then, try catching a ball while keeping your hands still. Hit water in a tub with your full palm. After the water has settled, hit the water again by diving your hand with your fingers first into the water. (Your full palm represents a swimmer doing a belly flop and your diving hand represents a swimmer doing a dive.) Explain what happens in each case and why. Which orientations would you advise people to avoid and why?

Note:

Making Connections: Constant Force and Constant Acceleration

The assumption of a constant force in the definition of impulse is analogous to the assumption of a constant acceleration in kinematics. In both cases, nature is adequately described without the use of calculus.

Section Summary

- Impulse, or change in momentum, equals the average net external force multiplied by the time this force acts:

Equation:

$$\Delta \mathbf{p} = \mathbf{F}_{\text{net}} \Delta t.$$

- Forces are usually not constant over a period of time.

Conceptual Questions

Exercise:

Problem: Professional Application

Explain in terms of impulse how padding reduces forces in a collision. State this in terms of a real example, such as the advantages of a carpeted vs. tile floor for a day care center.

Exercise:

Problem:

While jumping on a trampoline, sometimes you land on your back and other times on your feet. In which case can you reach a greater height and why?

Exercise:

Problem: Professional Application

Tennis racquets have “sweet spots.” If the ball hits a sweet spot then the player's arm is not jarred as much as it would be otherwise. Explain why this is the case.

Problems & Exercises

Exercise:

Problem:

A bullet is accelerated down the barrel of a gun by hot gases produced in the combustion of gun powder. What is the average force exerted on a 0.0300-kg bullet to accelerate it to a speed of 600 m/s in a time of 2.00 ms (milliseconds)?

Solution:

$$9.00 \times 10^3 \text{ N}$$

Exercise:**Problem: Professional Application**

A car moving at 10 m/s crashes into a tree and stops in 0.26 s. Calculate the force the seat belt exerts on a passenger in the car to bring him to a halt. The mass of the passenger is 70 kg.

Exercise:**Problem:**

A person slaps her leg with her hand, bringing her hand to rest in 2.50 milliseconds from an initial speed of 4.00 m/s. (a) What is the average force exerted on the leg, taking the effective mass of the hand and forearm to be 1.50 kg? (b) Would the force be any different if the woman clapped her hands together at the same speed and brought them to rest in the same time? Explain why or why not.

Solution:

a) $2.40 \times 10^3 \text{ N}$ toward the leg

b) The force on each hand would have the same magnitude as that found in part (a) (but in opposite directions by Newton's third law) because the change in momentum and the time interval are the same.

Exercise:

Problem: Professional Application

A professional boxer hits his opponent with a 1000-N horizontal blow that lasts for 0.150 s. (a) Calculate the impulse imparted by this blow. (b) What is the opponent's final velocity, if his mass is 105 kg and he is motionless in midair when struck near his center of mass? (c) Calculate the recoil velocity of the opponent's 10.0-kg head if hit in this manner, assuming the head does not initially transfer significant momentum to the boxer's body. (d) Discuss the implications of your answers for parts (b) and (c).

Exercise:**Problem: Professional Application**

Suppose a child drives a bumper car head on into the side rail, which exerts a force of 4000 N on the car for 0.200 s. (a) What impulse is imparted by this force? (b) Find the final velocity of the bumper car if its initial velocity was 2.80 m/s and the car plus driver have a mass of 200 kg. You may neglect friction between the car and floor.

Solution:

- a) 800 kg · m/s away from the wall
- b) 1.20 m/s away from the wall

Exercise:**Problem: Professional Application**

One hazard of space travel is debris left by previous missions. There are several thousand objects orbiting Earth that are large enough to be detected by radar, but there are far greater numbers of very small objects, such as flakes of paint. Calculate the force exerted by a 0.100-mg chip of paint that strikes a spacecraft window at a relative speed of 4.00×10^3 m/s, given the collision lasts 6.00×10^{-8} s.

Exercise:**Problem: Professional Application**

A 75.0-kg person is riding in a car moving at 20.0 m/s when the car runs into a bridge abutment. (a) Calculate the average force on the person if he is stopped by a padded dashboard that compresses an average of 1.00 cm. (b) Calculate the average force on the person if he is stopped by an air bag that compresses an average of 15.0 cm.

Solution:

(a) 1.50×10^6 N away from the dashboard

(b) 1.00×10^5 N away from the dashboard

Exercise:**Problem: Professional Application**

Military rifles have a mechanism for reducing the recoil forces of the gun on the person firing it. An internal part recoils over a relatively large distance and is stopped by damping mechanisms in the gun. The larger distance reduces the average force needed to stop the internal part. (a) Calculate the recoil velocity of a 1.00-kg plunger that directly interacts with a 0.0200-kg bullet fired at 600 m/s from the gun. (b) If this part is stopped over a distance of 20.0 cm, what average force is exerted upon it by the gun? (c) Compare this to the force exerted on the gun if the bullet is accelerated to its velocity in 10.0 ms (milliseconds).

Exercise:

Problem:

A cruise ship with a mass of 1.00×10^7 kg strikes a pier at a speed of 0.750 m/s. It comes to rest 6.00 m later, damaging the ship, the pier, and the tugboat captain's finances. Calculate the average force exerted on the pier using the concept of impulse. (Hint: First calculate the time it took to bring the ship to rest.)

Solution:

4.69×10^5 N in the boat's original direction of motion

Exercise:**Problem:**

Calculate the final speed of a 110-kg rugby player who is initially running at 8.00 m/s but collides head-on with a padded goalpost and experiences a backward force of 1.76×10^4 N for 5.50×10^{-2} s.

Exercise:**Problem:**

Water from a fire hose is directed horizontally against a wall at a rate of 50.0 kg/s and a speed of 42.0 m/s. Calculate the magnitude of the force exerted on the wall, assuming the water's horizontal momentum is reduced to zero.

Solution:

2.10×10^3 N away from the wall

Exercise:**Problem:**

A 0.450-kg hammer is moving horizontally at 7.00 m/s when it strikes a nail and comes to rest after driving the nail 1.00 cm into a board. (a) Calculate the duration of the impact. (b) What was the average force exerted on the nail?

Exercise:**Problem:**

Starting with the definitions of momentum and kinetic energy, derive an equation for the kinetic energy of a particle expressed as a function of its momentum.

Solution:**Equation:**

$$\begin{aligned}\mathbf{p} &= m\mathbf{v} \Rightarrow p^2 = m^2v^2 \Rightarrow \frac{p^2}{m} = mv^2 \\ \Rightarrow \frac{p^2}{2m} &= \frac{1}{2}mv^2 = \text{KE} \\ KE &= \frac{p^2}{2m}\end{aligned}$$

Exercise:**Problem:**

A ball with an initial velocity of 10 m/s moves at an angle 60° above the $+x$ -direction. The ball hits a vertical wall and bounces off so that it is moving 60° above the $-x$ -direction with the same speed. What is the impulse delivered by the wall?

Exercise:**Problem:**

When serving a tennis ball, a player hits the ball when its velocity is zero (at the highest point of a vertical toss). The racquet exerts a force of 540 N on the ball for 5.00 ms, giving it a final velocity of 45.0 m/s. Using these data, find the mass of the ball.

Solution:

60.0 g

Exercise:

Problem:

A punter drops a ball from rest vertically 1 meter down onto his foot. The ball leaves the foot with a speed of 18 m/s at an angle 55° above the horizontal. What is the impulse delivered by the foot (magnitude and direction)?

Glossary

change in momentum

the difference between the final and initial momentum; the mass times the change in velocity

impulse

the average net external force times the time it acts; equal to the change in momentum

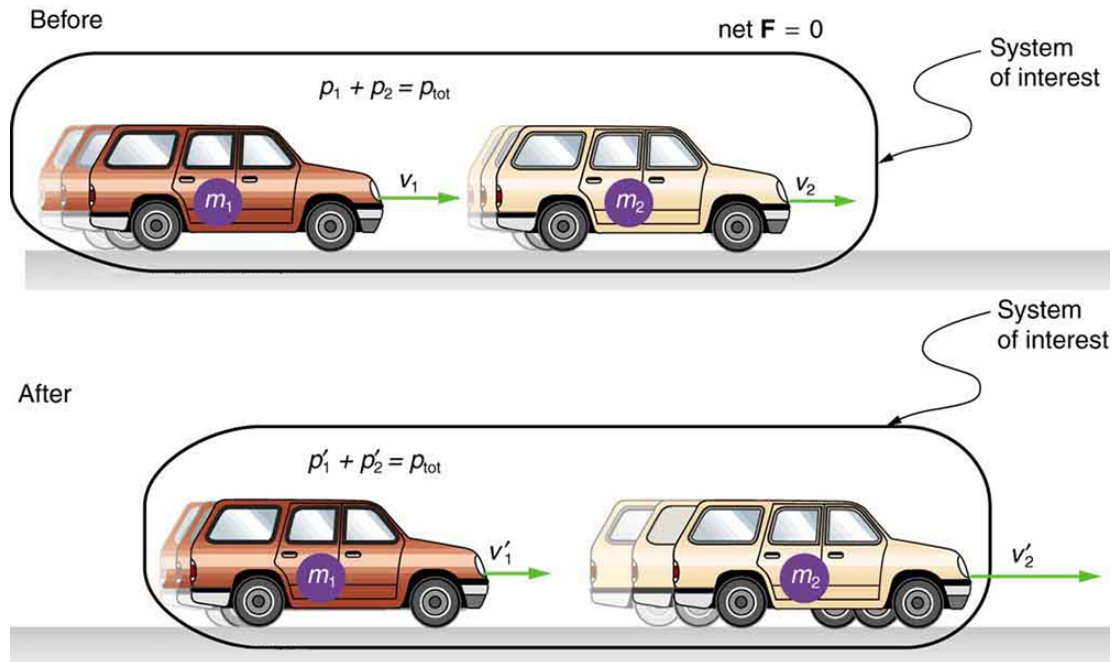
Conservation of Momentum

- Describe the principle of conservation of momentum.
- Derive an expression for the conservation of momentum.
- Explain conservation of momentum with examples.
- Explain the principle of conservation of momentum as it relates to atomic and subatomic particles.

Momentum is an important quantity because it is conserved. Yet it was not conserved in the examples in [Impulse](#) and [Linear Momentum and Force](#), where large changes in momentum were produced by forces acting on the system of interest. Under what circumstances is momentum conserved?

The answer to this question entails considering a sufficiently large system. It is always possible to find a larger system in which total momentum is constant, even if momentum changes for components of the system. If a football player runs into the goalpost in the end zone, there will be a force on him that causes him to bounce backward. However, the Earth also recoils—conserving momentum—because of the force applied to it through the goalpost. Because Earth is many orders of magnitude more massive than the player, its recoil is immeasurably small and can be neglected in any practical sense, but it is real nevertheless.

Consider what happens if the masses of two colliding objects are more similar than the masses of a football player and Earth—for example, one car bumping into another, as shown in [\[link\]](#). Both cars are coasting in the same direction when the lead car (labeled m_2) is bumped by the trailing car (labeled m_1). The only unbalanced force on each car is the force of the collision. (Assume that the effects due to friction are negligible.) Car 1 slows down as a result of the collision, losing some momentum, while car 2 speeds up and gains some momentum. We shall now show that the total momentum of the two-car system remains constant.



A car of mass m_1 moving with a velocity of v_1 bumps into another car of mass m_2 and velocity v_2 that it is following. As a result, the first car slows down to a velocity of v'_1 and the second speeds up to a velocity of v'_2 . The momentum of each car is changed, but the total momentum p_{tot} of the two cars is the same before and after the collision (if you assume friction is negligible).

Using the definition of impulse, the change in momentum of car 1 is given by

Equation:

$$\Delta p_1 = F_1 \Delta t,$$

where F_1 is the force on car 1 due to car 2, and Δt is the time the force acts (the duration of the collision). Intuitively, it seems obvious that the collision time is the same for both cars, but it is only true for objects traveling at ordinary speeds. This assumption must be modified for objects travelling

near the speed of light, without affecting the result that momentum is conserved.

Similarly, the change in momentum of car 2 is

Equation:

$$\Delta p_2 = F_2 \Delta t,$$

where F_2 is the force on car 2 due to car 1, and we assume the duration of the collision Δt is the same for both cars. We know from Newton's third law that $F_2 = -F_1$, and so

Equation:

$$\Delta p_2 = -F_1 \Delta t = -\Delta p_1.$$

Thus, the changes in momentum are equal and opposite, and

Equation:

$$\Delta p_1 + \Delta p_2 = 0.$$

Because the changes in momentum add to zero, the total momentum of the two-car system is constant. That is,

Equation:

$$p_1 + p_2 = \text{constant},$$

Equation:

$$p_1 + p_2 = p'_1 + p'_2,$$

where p'_1 and p'_2 are the momenta of cars 1 and 2 after the collision. (We often use primes to denote the final state.)

This result—that momentum is conserved—has validity far beyond the preceding one-dimensional case. It can be similarly shown that total momentum is conserved for any isolated system, with any number of

objects in it. In equation form, the **conservation of momentum principle** for an isolated system is written

Equation:

$$\mathbf{p}_{\text{tot}} = \text{constant},$$

or

Equation:

$$\mathbf{p}_{\text{tot}} = \mathbf{p}'_{\text{tot}},$$

where \mathbf{p}_{tot} is the total momentum (the sum of the momenta of the individual objects in the system) and \mathbf{p}'_{tot} is the total momentum some time later. (The total momentum can be shown to be the momentum of the center of mass of the system.) An **isolated system** is defined to be one for which the net external force is zero ($\mathbf{F}_{\text{net}} = 0$).

Note:

Conservation of Momentum Principle

Equation:

$$\begin{aligned}\mathbf{p}_{\text{tot}} &= \text{constant} \\ \mathbf{p}_{\text{tot}} &= \mathbf{p}'_{\text{tot}} \text{ (isolated system)}\end{aligned}$$

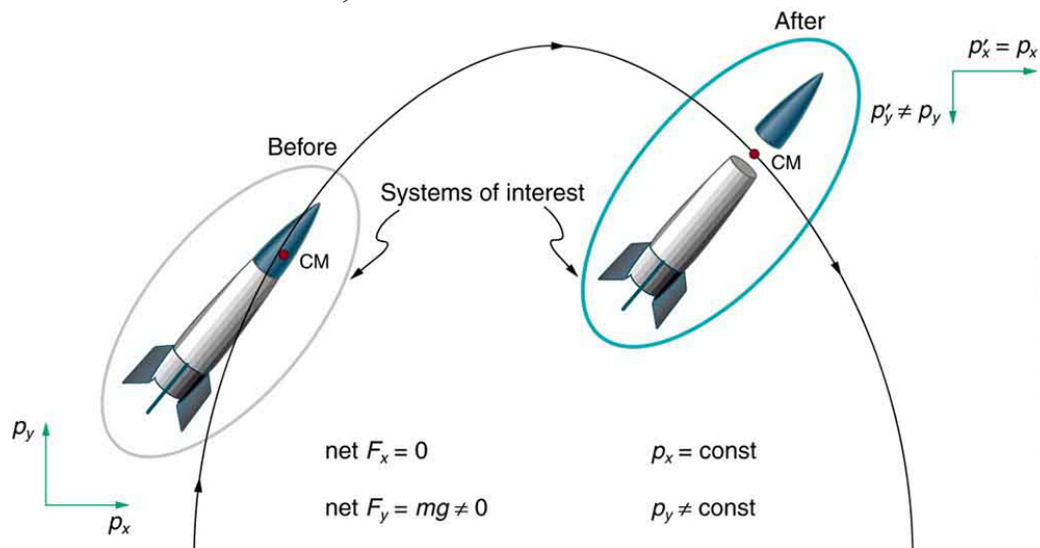
Note:

Isolated System

An isolated system is defined to be one for which the net external force is zero ($\mathbf{F}_{\text{net}} = 0$).

Perhaps an easier way to see that momentum is conserved for an isolated system is to consider Newton's second law in terms of momentum, $\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}_{\text{tot}}}{\Delta t}$. For an isolated system, ($\mathbf{F}_{\text{net}} = 0$); thus, $\Delta \mathbf{p}_{\text{tot}} = 0$, and \mathbf{p}_{tot} is constant.

We have noted that the three length dimensions in nature— x , y , and z —are independent, and it is interesting to note that momentum can be conserved in different ways along each dimension. For example, during projectile motion and where air resistance is negligible, momentum is conserved in the horizontal direction because horizontal forces are zero and momentum is unchanged. But along the vertical direction, the net vertical force is not zero and the momentum of the projectile is not conserved. (See [\[link\]](#).) However, if the momentum of the projectile-Earth system is considered in the vertical direction, we find that the total momentum is conserved.



The horizontal component of a projectile's momentum is conserved if air resistance is negligible, even in this case where a space probe separates. The forces causing the separation are internal to the system, so that the net external horizontal force $F_{x-\text{net}}$ is still zero. The vertical component of the momentum is not conserved, because the net vertical force $F_{y-\text{net}}$ is not zero. In the vertical direction, the space probe-Earth system needs to be considered and we find that the total momentum is conserved. The center of mass of the

space probe takes the same path it would if the separation did not occur.

The conservation of momentum principle can be applied to systems as different as a comet striking Earth and a gas containing huge numbers of atoms and molecules. Conservation of momentum is violated only when the net external force is not zero. But another larger system can always be considered in which momentum is conserved by simply including the source of the external force. For example, in the collision of two cars considered above, the two-car system conserves momentum while each one-car system does not.

Note:

Making Connections: Take-Home Investigation—Drop of Tennis Ball and a Basketball

Hold a tennis ball side by side and in contact with a basketball. Drop the balls together. (Be careful!) What happens? Explain your observations. Now hold the tennis ball above and in contact with the basketball. What happened? Explain your observations. What do you think will happen if the basketball ball is held above and in contact with the tennis ball?

Note:

Making Connections: Take-Home Investigation—Two Tennis Balls in a Ballistic Trajectory

Tie two tennis balls together with a string about a foot long. Hold one ball and let the other hang down and throw it in a ballistic trajectory. Explain your observations. Now mark the center of the string with bright ink or attach a brightly colored sticker to it and throw again. What happened? Explain your observations.

Some aquatic animals such as jellyfish move around based on the principles of conservation of momentum. A jellyfish fills its umbrella section with water and then pushes the water out resulting in motion in the opposite direction to that of the jet of water. Squids propel themselves in a

similar manner but, in contrast with jellyfish, are able to control the direction in which they move by aiming their nozzle forward or backward. Typical squids can move at speeds of 8 to 12 km/h.

The ballistocardiograph (BCG) was a diagnostic tool used in the second half of the 20th century to study the strength of the heart. About once a second, your heart beats, forcing blood into the aorta. A force in the opposite direction is exerted on the rest of your body (recall Newton's third law). A ballistocardiograph is a device that can measure this reaction force. This measurement is done by using a sensor (resting on the person) or by using a moving table suspended from the ceiling. This technique can gather information on the strength of the heart beat and the volume of blood passing from the heart. However, the electrocardiogram (ECG or EKG) and the echocardiogram (cardiac ECHO or ECHO; a technique that uses ultrasound to see an image of the heart) are more widely used in the practice of cardiology.

Note:

Making Connections: Conservation of Momentum and Collision

Conservation of momentum is quite useful in describing collisions. Momentum is crucial to our understanding of atomic and subatomic particles because much of what we know about these particles comes from collision experiments.

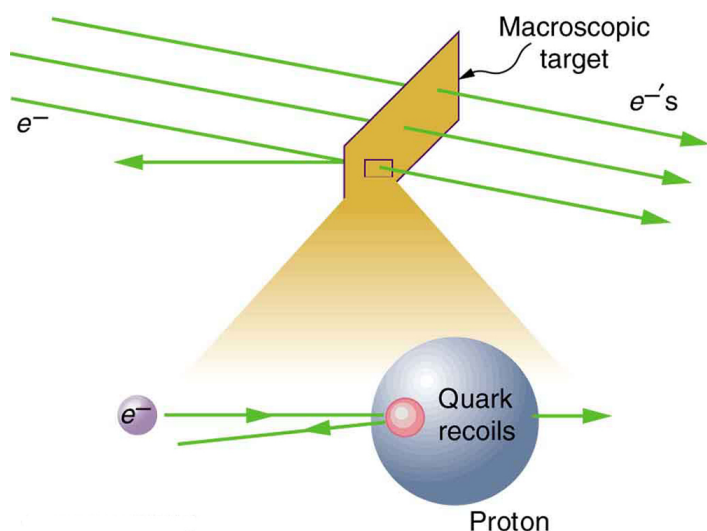
Subatomic Collisions and Momentum

The conservation of momentum principle not only applies to the macroscopic objects, it is also essential to our explorations of atomic and subatomic particles. Giant machines hurl subatomic particles at one another, and researchers evaluate the results by assuming conservation of momentum (among other things).

On the small scale, we find that particles and their properties are invisible to the naked eye but can be measured with our instruments, and models of these subatomic particles can be constructed to describe the results.

Momentum is found to be a property of all subatomic particles including massless particles such as photons that compose light. Momentum being a property of particles hints that momentum may have an identity beyond the description of an object's mass multiplied by the object's velocity. Indeed, momentum relates to wave properties and plays a fundamental role in what measurements are taken and how we take these measurements.

Furthermore, we find that the conservation of momentum principle is valid when considering systems of particles. We use this principle to analyze the masses and other properties of previously undetected particles, such as the nucleus of an atom and the existence of quarks that make up particles of nuclei. [\[link\]](#) below illustrates how a particle scattering backward from another implies that its target is massive and dense. Experiments seeking evidence that **quarks** make up protons (one type of particle that makes up nuclei) scattered high-energy electrons off of protons (nuclei of hydrogen atoms). Electrons occasionally scattered straight backward in a manner that implied a very small and very dense particle makes up the proton—this observation is considered nearly direct evidence of quarks. The analysis was based partly on the same conservation of momentum principle that works so well on the large scale.



A subatomic particle scatters straight backward from a target particle. In experiments seeking evidence for

quarks, electrons were observed to occasionally scatter straight backward from a proton.

Section Summary

- The conservation of momentum principle is written
Equation:

$$\mathbf{p}_{\text{tot}} = \text{constant}$$

or

Equation:

$$\mathbf{p}_{\text{tot}} = \mathbf{p}'_{\text{tot}} \text{ (isolated system),}$$

\mathbf{p}_{tot} is the initial total momentum and \mathbf{p}'_{tot} is the total momentum some time later.

- An isolated system is defined to be one for which the net external force is zero ($\mathbf{F}_{\text{net}} = 0$).
- During projectile motion and where air resistance is negligible, momentum is conserved in the horizontal direction because horizontal forces are zero.
- Conservation of momentum applies only when the net external force is zero.
- The conservation of momentum principle is valid when considering systems of particles.

Conceptual Questions

Exercise:

Problem: Professional Application

If you dive into water, you reach greater depths than if you do a belly flop. Explain this difference in depth using the concept of conservation of energy. Explain this difference in depth using what you have learned in this chapter.

Exercise:

Problem: Under what circumstances is momentum conserved?

Exercise:

Problem:

Can momentum be conserved for a system if there are external forces acting on the system? If so, under what conditions? If not, why not?

Exercise:

Problem:

Momentum for a system can be conserved in one direction while not being conserved in another. What is the angle between the directions? Give an example.

Exercise:

Problem: Professional Application

Explain in terms of momentum and Newton's laws how a car's air resistance is due in part to the fact that it pushes air in its direction of motion.

Exercise:

Problem:

Can objects in a system have momentum while the momentum of the system is zero? Explain your answer.

Exercise:

Problem:

Must the total energy of a system be conserved whenever its momentum is conserved? Explain why or why not.

Problems & Exercises**Exercise:****Problem: Professional Application**

Train cars are coupled together by being bumped into one another. Suppose two loaded train cars are moving toward one another, the first having a mass of 150,000 kg and a velocity of 0.300 m/s, and the second having a mass of 110,000 kg and a velocity of -0.120 m/s. (The minus indicates direction of motion.) What is their final velocity?

Solution:

0.122 m/s

Exercise:**Problem:**

Suppose a clay model of a koala bear has a mass of 0.200 kg and slides on ice at a speed of 0.750 m/s. It runs into another clay model, which is initially motionless and has a mass of 0.350 kg. Both being soft clay, they naturally stick together. What is their final velocity?

Exercise:**Problem: Professional Application**

Consider the following question: *A car moving at 10 m/s crashes into a tree and stops in 0.26 s. Calculate the force the seatbelt exerts on a passenger in the car to bring him to a halt. The mass of the passenger is 70 kg.* Would the answer to this question be different if the car with

the 70-kg passenger had collided with a car that has a mass equal to and is traveling in the opposite direction and at the same speed? Explain your answer.

Solution:

In a collision with an identical car, momentum is conserved. Afterwards $v_f = 0$ for both cars. The change in momentum will be the same as in the crash with the tree. However, the force on the body is not determined since the time is not known. A padded stop will reduce injurious force on body.

Exercise:

Problem:

What is the velocity of a 900-kg car initially moving at 30.0 m/s, just after it hits a 150-kg deer initially running at 12.0 m/s in the same direction? Assume the deer remains on the car.

Exercise:

Problem:

A 1.80-kg falcon catches a 0.650-kg dove from behind in midair. What is their velocity after impact if the falcon's velocity is initially 28.0 m/s and the dove's velocity is 7.00 m/s in the same direction?

Solution:

22.4 m/s in the same direction as the original motion

Glossary

conservation of momentum principle

when the net external force is zero, the total momentum of the system is conserved or constant

isolated system

a system in which the net external force is zero

quark

fundamental constituent of matter and an elementary particle

Elastic Collisions in One Dimension

- Describe an elastic collision of two objects in one dimension.
- Define internal kinetic energy.
- Derive an expression for conservation of internal kinetic energy in a one dimensional collision.
- Determine the final velocities in an elastic collision given masses and initial velocities.

Let us consider various types of two-object collisions. These collisions are the easiest to analyze, and they illustrate many of the physical principles involved in collisions. The conservation of momentum principle is very useful here, and it can be used whenever the net external force on a system is zero.

We start with the elastic collision of two objects moving along the same line—a one-dimensional problem. An **elastic collision** is one that also conserves internal kinetic energy. **Internal kinetic energy** is the sum of the kinetic energies of the objects in the system. [\[link\]](#) illustrates an elastic collision in which internal kinetic energy and momentum are conserved.

Truly elastic collisions can only be achieved with subatomic particles, such as electrons striking nuclei. Macroscopic collisions can be very nearly, but not quite, elastic—some kinetic energy is always converted into other forms of energy such as heat transfer due to friction and sound. One macroscopic collision that is nearly elastic is that of two steel blocks on ice. Another nearly elastic collision is that between two carts with spring bumpers on an air track. Icy surfaces and air tracks are nearly frictionless, more readily allowing nearly elastic collisions on them.

Note:

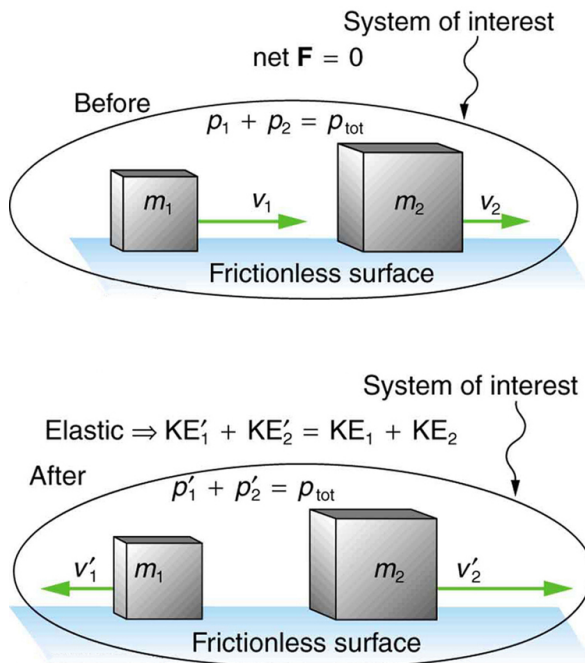
Elastic Collision

An **elastic collision** is one that conserves internal kinetic energy.

Note:

Internal Kinetic Energy

Internal kinetic energy is the sum of the kinetic energies of the objects in the system.



An elastic one-dimensional
two-object collision.
Momentum and internal kinetic
energy are conserved.

Now, to solve problems involving one-dimensional elastic collisions between two objects we can use the equations for conservation of momentum and conservation of internal kinetic energy. First, the equation for conservation of momentum for two objects in a one-dimensional collision is

Equation:

$$p_1 + p_2 = p'_1 + p'_2 \quad (F_{\text{net}} = 0)$$

or

Equation:

$$m_1v_1 + m_2v_2 = m_1v'_1 + m_2v'_2 \quad (F_{\text{net}} = 0),$$

where the primes (') indicate values after the collision. By definition, an elastic collision conserves internal kinetic energy, and so the sum of kinetic energies before the collision equals the sum after the collision. Thus,

Equation:

$$\frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 = \frac{1}{2}m_1v'^2_1 + \frac{1}{2}m_2v'^2_2 \quad (\text{two-object elastic collision})$$

expresses the equation for conservation of internal kinetic energy in a one-dimensional collision.

Example:**Calculating Velocities Following an Elastic Collision**

Calculate the velocities of two objects following an elastic collision, given that

Equation:

$$m_1 = 0.500 \text{ kg}, \quad m_2 = 3.50 \text{ kg}, \quad v_1 = 4.00 \text{ m/s}, \text{ and } v_2 = 0.$$

Strategy and Concept

First, visualize what the initial conditions mean—a small object strikes a larger object that is initially at rest. This situation is slightly simpler than the situation shown in [\[link\]](#) where both objects are initially moving. We are asked to find two unknowns (the final velocities v'_1 and v'_2). To find two unknowns, we must use two independent equations. Because this collision is elastic, we can use the above two equations. Both can be simplified by the fact that object 2 is initially at rest, and thus $v_2 = 0$. Once we simplify these equations, we combine them algebraically to solve for the unknowns.

Solution

For this problem, note that $v_2 = 0$ and use conservation of momentum. Thus,

Equation:

$$p_1 = p'_1 + p'_2$$

or

Equation:

$$m_1 v_1 = m_1 v'_1 + m_2 v'_2.$$

Using conservation of internal kinetic energy and that $v_2 = 0$,

Equation:

$$\frac{1}{2} m_1 v_1^2 = \frac{1}{2} m_1 v'^2_1 + \frac{1}{2} m_2 v'^2_2.$$

Solving the first equation (momentum equation) for v'_2 , we obtain

Equation:

$$v'_2 = \frac{m_1}{m_2} (v_1 - v'_1).$$

Substituting this expression into the second equation (internal kinetic energy equation) eliminates the variable v'_2 , leaving only v'_1 as an unknown (the algebra is left as an exercise for the reader). There are two solutions to any quadratic equation; in this example, they are

Equation:

$$v'_1 = 4.00 \text{ m/s}$$

and

Equation:

$$v'_1 = -3.00 \text{ m/s}.$$

As noted when quadratic equations were encountered in earlier chapters, both solutions may or may not be meaningful. In this case, the first solution is the same as the initial condition. The first solution thus represents the situation before the collision and is discarded. The second solution ($v'_1 = -3.00 \text{ m/s}$) is negative, meaning that the first object bounces backward. When this negative value of v'_1 is used to find the velocity of the second object after the collision, we get

Equation:

$$v'_2 = \frac{m_1}{m_2}(v_1 - v'_1) = \frac{0.500 \text{ kg}}{3.50 \text{ kg}}[4.00 - (-3.00)] \text{ m/s}$$

or

Equation:

$$v'_2 = 1.00 \text{ m/s.}$$

Discussion

The result of this example is intuitively reasonable. A small object strikes a larger one at rest and bounces backward. The larger one is knocked forward, but with a low speed. (This is like a compact car bouncing backward off a full-size SUV that is initially at rest.) As a check, try calculating the internal kinetic energy before and after the collision. You will see that the internal kinetic energy is unchanged at 4.00 J. Also check the total momentum before and after the collision; you will find it, too, is unchanged.

The equations for conservation of momentum and internal kinetic energy as written above can be used to describe any one-dimensional elastic collision of two objects. These equations can be extended to more objects if needed.

Note:

Making Connections: Take-Home Investigation—Ice Cubes and Elastic Collision

Find a few ice cubes which are about the same size and a smooth kitchen tabletop or a table with a glass top. Place the ice cubes on the surface several centimeters away from each other. Flick one ice cube toward a stationary ice cube and observe the path and velocities of the ice cubes after the collision. Try to avoid edge-on collisions and collisions with rotating ice cubes. Have you created approximately elastic collisions? Explain the speeds and directions of the ice cubes using momentum.

Note:

PhET Explorations: Collision Lab

Investigate collisions on an air hockey table. Set up your own experiments: vary the number of discs, masses and initial conditions. Is momentum

conserved? Is kinetic energy conserved? Vary the elasticity and see what happens.

https://phet.colorado.edu/sims/collision-lab/collision-lab_en.html

Section Summary

- An elastic collision is one that conserves internal kinetic energy.
- Conservation of kinetic energy and momentum together allow the final velocities to be calculated in terms of initial velocities and masses in one dimensional two-body collisions.

Conceptual Questions

Exercise:

Problem: What is an elastic collision?

Problems & Exercises

Exercise:

Problem:

Two identical objects (such as billiard balls) have a one-dimensional collision in which one is initially motionless. After the collision, the moving object is stationary and the other moves with the same speed as the other originally had. Show that both momentum and kinetic energy are conserved.

Exercise:

Problem: Professional Application

Two manned satellites approach one another at a relative speed of 0.250 m/s, intending to dock. The first has a mass of 4.00×10^3 kg, and the

second a mass of 7.50×10^3 kg. If the two satellites collide elastically rather than dock, what is their final relative velocity?

Solution:

0.250 m/s

Exercise:

Problem:

A 70.0-kg ice hockey goalie, originally at rest, catches a 0.150-kg hockey puck slapped at him at a velocity of 35.0 m/s. Suppose the goalie and the ice puck have an elastic collision and the puck is reflected back in the direction from which it came. What would their final velocities be in this case?

Glossary

elastic collision

a collision that also conserves internal kinetic energy

internal kinetic energy

the sum of the kinetic energies of the objects in a system

Inelastic Collisions in One Dimension

- Define inelastic collision.
- Explain perfectly inelastic collision.
- Apply an understanding of collisions to sports.
- Determine recoil velocity and loss in kinetic energy given mass and initial velocity.

We have seen that in an elastic collision, internal kinetic energy is conserved. An **inelastic collision** is one in which the internal kinetic energy changes (it is not conserved). This lack of conservation means that the forces between colliding objects may remove or add internal kinetic energy. Work done by internal forces may change the forms of energy within a system. For inelastic collisions, such as when colliding objects stick together, this internal work may transform some internal kinetic energy into heat transfer. Or it may convert stored energy into internal kinetic energy, such as when exploding bolts separate a satellite from its launch vehicle.

Note:

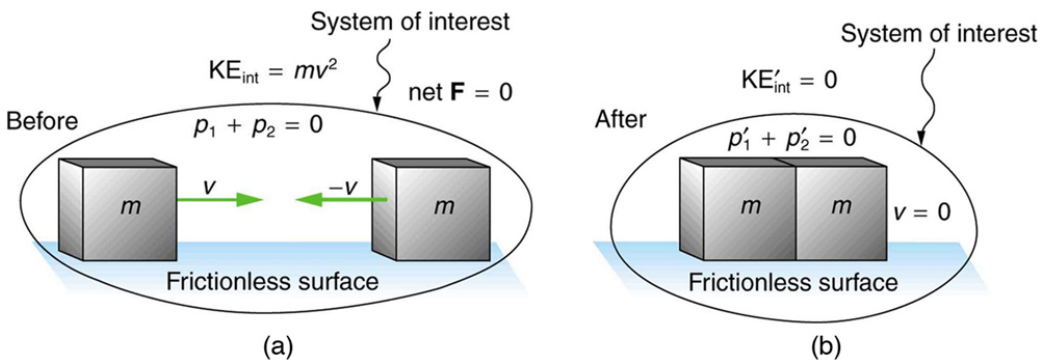
Inelastic Collision

An inelastic collision is one in which the internal kinetic energy changes (it is not conserved).

[\[link\]](#) shows an example of an inelastic collision. Two objects that have equal masses head toward one another at equal speeds and then stick together. Their total internal kinetic energy is initially $\frac{1}{2}mv^2 + \frac{1}{2}mv^2 = mv^2$. The two objects come to rest after sticking together, conserving momentum. But the internal kinetic energy is zero after the collision. A collision in which the objects stick together is sometimes called a **perfectly inelastic collision** because it reduces internal kinetic energy more than does any other type of inelastic collision. In fact, such a collision reduces internal kinetic energy to the minimum it can have while still conserving momentum.

Note:**Perfectly Inelastic Collision**

A collision in which the objects stick together is sometimes called “perfectly inelastic.”



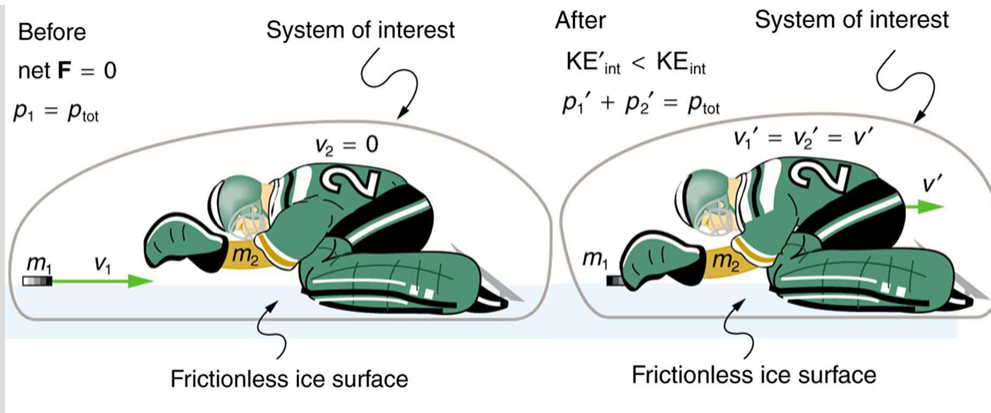
An inelastic one-dimensional two-object collision. Momentum is conserved, but internal kinetic energy is not conserved. (a) Two objects of equal mass initially head directly toward one another at the same speed. (b) The objects stick together (a perfectly inelastic collision), and so their final velocity is zero. The internal kinetic energy of the system changes in any inelastic collision and is reduced to zero in this example.

Example:**Calculating Velocity and Change in Kinetic Energy: Inelastic Collision of a Puck and a Goalie**

(a) Find the recoil velocity of a 70.0-kg ice hockey goalie, originally at rest, who catches a 0.150-kg hockey puck slapped at him at a velocity of 35.0 m/s. (b) How much kinetic energy is lost during the collision?

Assume friction between the ice and the puck-goalie system is negligible.

(See [link](#))



An ice hockey goalie catches a hockey puck and recoils backward. The initial kinetic energy of the puck is almost entirely converted to thermal energy and sound in this inelastic collision.

Strategy

Momentum is conserved because the net external force on the puck-goalie system is zero. We can thus use conservation of momentum to find the final velocity of the puck and goalie system. Note that the initial velocity of the goalie is zero and that the final velocity of the puck and goalie are the same. Once the final velocity is found, the kinetic energies can be calculated before and after the collision and compared as requested.

Solution for (a)

Momentum is conserved because the net external force on the puck-goalie system is zero.

Conservation of momentum is

Equation:

$$p_1 + p_2 = p'_1 + p'_2$$

or

Equation:

$$m_1 v_1 + m_2 v_2 = m_1 v'_1 + m_2 v'_2.$$

Because the goalie is initially at rest, we know $v_2 = 0$. Because the goalie catches the puck, the final velocities are equal, or $v'_1 = v'_2 = v'$. Thus, the

conservation of momentum equation simplifies to

Equation:

$$m_1 v_1 = (m_1 + m_2) v_f.$$

Solving for v_f yields

Equation:

$$v_f = \frac{m_1}{m_1 + m_2} v_1.$$

Entering known values in this equation, we get

Equation:

$$v_f = \left(\frac{0.150 \text{ kg}}{0.150 \text{ kg} + 70.0 \text{ kg}} \right) (35.0 \text{ m/s}) = 7.48 \times 10^{-2} \text{ m/s}.$$

Discussion for (a)

This recoil velocity is small and in the same direction as the puck's original velocity, as we might expect.

Solution for (b)

Before the collision, the internal kinetic energy KE_{int} of the system is that of the hockey puck, because the goalie is initially at rest. Therefore, KE_{int} is initially

Equation:

$$\begin{aligned} \text{KE}_{\text{int}} &= \frac{1}{2} m v^2 = \frac{1}{2} (0.150 \text{ kg}) (35.0 \text{ m/s})^2 \\ &= 91.9 \text{ J}. \end{aligned}$$

After the collision, the internal kinetic energy is

Equation:

$$\begin{aligned} \text{KE}'_{\text{int}} &= \frac{1}{2} (m + M) v^2 = \frac{1}{2} (70.15 \text{ kg}) (7.48 \times 10^{-2} \text{ m/s})^2 \\ &= 0.196 \text{ J}. \end{aligned}$$

The change in internal kinetic energy is thus

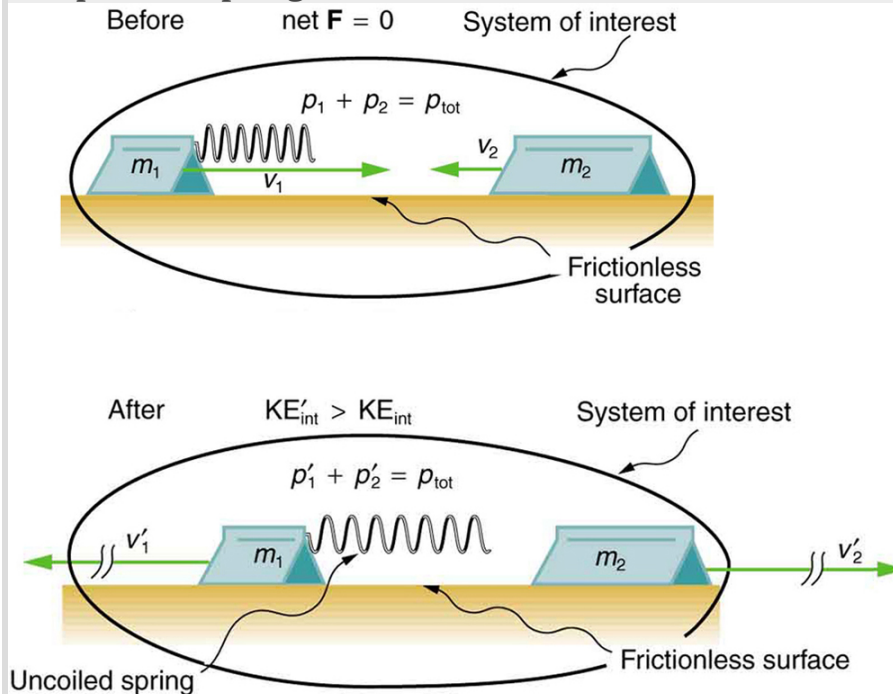
Equation:

$$\begin{aligned} KE_{\text{int}} - KE_{\text{int}} &= 0.196 \text{ J} - 91.9 \text{ J} \\ &= -91.7 \text{ J} \end{aligned}$$

where the minus sign indicates that the energy was lost.

Discussion for (b)

Nearly all of the initial internal kinetic energy is lost in this perfectly inelastic collision. KE_{int} is mostly converted to thermal energy and sound. During some collisions, the objects do not stick together and less of the internal kinetic energy is removed—such as happens in most automobile accidents. Alternatively, stored energy may be converted into internal kinetic energy during a collision. [\[link\]](#) shows a one-dimensional example in which two carts on an air track collide, releasing potential energy from a compressed spring. [\[link\]](#) deals with data from such a collision.



An air track is nearly frictionless, so that momentum is conserved. Motion is one-dimensional. In this collision, examined in [\[link\]](#), the potential energy of a compressed spring is released during the collision and is converted to internal kinetic energy.

Collisions are particularly important in sports and the sporting and leisure industry utilizes elastic and inelastic collisions. Let us look briefly at tennis. Recall that in a collision, it is momentum and not force that is important. So, a heavier tennis racquet will have the advantage over a lighter one. This conclusion also holds true for other sports—a lightweight bat (such as a softball bat) cannot hit a hardball very far.

The location of the impact of the tennis ball on the racquet is also important, as is the part of the stroke during which the impact occurs. A smooth motion results in the maximizing of the velocity of the ball after impact and reduces sports injuries such as tennis elbow. A tennis player tries to hit the ball on the “sweet spot” on the racquet, where the vibration and impact are minimized and the ball is able to be given more velocity. Sports science and technologies also use physics concepts such as momentum and rotational motion and vibrations.

Note:**Take-Home Experiment—Bouncing of Tennis Ball**

1. Find a racquet (a tennis, badminton, or other racquet will do). Place the racquet on the floor and stand on the handle. Drop a tennis ball on the strings from a measured height. Measure how high the ball bounces. Now ask a friend to hold the racquet firmly by the handle and drop a tennis ball from the same measured height above the racquet. Measure how high the ball bounces and observe what happens to your friend’s hand during the collision. Explain your observations and measurements.
2. The coefficient of restitution (c) is a measure of the elasticity of a collision between a ball and an object, and is defined as the ratio of the speeds after and before the collision. A perfectly elastic collision has a c of 1. For a ball bouncing off the floor (or a racquet on the floor), c can be shown to be $c = (h/H)^{1/2}$ where h is the height to which the ball bounces and H is the height from which the ball is dropped. Determine c for the cases in Part 1 and for the case of a tennis ball bouncing off a concrete or wooden floor ($c = 0.85$ for new tennis balls used on a tennis court).

Example:**Calculating Final Velocity and Energy Release: Two Carts Collide**

In the collision pictured in [\[link\]](#), two carts collide inelastically. Cart 1 (denoted m_1) carries a spring which is initially compressed. During the collision, the spring releases its potential energy and converts it to internal kinetic energy. The mass of cart 1 and the spring is 0.350 kg, and the cart and the spring together have an initial velocity of 2.00 m/s. Cart 2 (denoted m_2 in [\[link\]](#)) has a mass of 0.500 kg and an initial velocity of -0.500 m/s. After the collision, cart 1 is observed to recoil with a velocity of -4.00 m/s. (a) What is the final velocity of cart 2? (b) How much energy was released by the spring (assuming all of it was converted into internal kinetic energy)?

Strategy

We can use conservation of momentum to find the final velocity of cart 2, because $F_{\text{net}} = 0$ (the track is frictionless and the force of the spring is internal). Once this velocity is determined, we can compare the internal kinetic energy before and after the collision to see how much energy was released by the spring.

Solution for (a)

As before, the equation for conservation of momentum in a two-object system is

Equation:

$$m_1 v_1 + m_2 v_2 = m_1 v'_1 + m_2 v'_2.$$

The only unknown in this equation is v'_2 . Solving for v'_2 and substituting known values into the previous equation yields

Equation:

$$\begin{aligned} v'_2 &= \frac{m_1 v_1 + m_2 v_2 - m_1 v'_1}{m_2} \\ &= \frac{(0.350 \text{ kg})(2.00 \text{ m/s}) + (0.500 \text{ kg})(-0.500 \text{ m/s})}{0.500 \text{ kg}} - \frac{(0.350 \text{ kg})(-4.00 \text{ m/s})}{0.500 \text{ kg}} \\ &= 3.70 \text{ m/s.} \end{aligned}$$

Solution for (b)

The internal kinetic energy before the collision is

Equation:

$$\begin{aligned}\text{KE}_{\text{int}} &= \frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 \\ &= \frac{1}{2}(0.350 \text{ kg})(2.00 \text{ m/s})^2 + \frac{1}{2}(0.500 \text{ kg})(-0.500 \text{ m/s})^2 \\ &= 0.763 \text{ J}.\end{aligned}$$

After the collision, the internal kinetic energy is

Equation:

$$\begin{aligned}\text{KE}'_{\text{int}} &= \frac{1}{2}m_1v_1'^2 + \frac{1}{2}m_2v_2'^2 \\ &= \frac{1}{2}(0.350 \text{ kg})(-4.00 \text{ m/s})^2 + \frac{1}{2}(0.500 \text{ kg})(3.70 \text{ m/s})^2 \\ &= 6.22 \text{ J}.\end{aligned}$$

The change in internal kinetic energy is thus

Equation:

$$\begin{aligned}\text{KE}'_{\text{int}} - \text{KE}_{\text{int}} &= 6.22 \text{ J} - 0.763 \text{ J} \\ &= 5.46 \text{ J}.\end{aligned}$$

Discussion

The final velocity of cart 2 is large and positive, meaning that it is moving to the right after the collision. The internal kinetic energy in this collision increases by 5.46 J. That energy was released by the spring.

Section Summary

- An inelastic collision is one in which the internal kinetic energy changes (it is not conserved).
- A collision in which the objects stick together is sometimes called perfectly inelastic because it reduces internal kinetic energy more than does any other type of inelastic collision.

- Sports science and technologies also use physics concepts such as momentum and rotational motion and vibrations.

Conceptual Questions

Exercise:

Problem:

What is an inelastic collision? What is a perfectly inelastic collision?

Exercise:

Problem:

Mixed-pair ice skaters performing in a show are standing motionless at arms length just before starting a routine. They reach out, clasp hands, and pull themselves together by only using their arms. Assuming there is no friction between the blades of their skates and the ice, what is their velocity after their bodies meet?

Exercise:

Problem:

A small pickup truck that has a camper shell slowly coasts toward a red light with negligible friction. Two dogs in the back of the truck are moving and making various inelastic collisions with each other and the walls. What is the effect of the dogs on the motion of the center of mass of the system (truck plus entire load)? What is their effect on the motion of the truck?

Problems & Exercises

Exercise:

Problem:

A 0.240-kg billiard ball that is moving at 3.00 m/s strikes the bumper of a pool table and bounces straight back at 2.40 m/s (80% of its original speed). The collision lasts 0.0150 s. (a) Calculate the average force exerted on the ball by the bumper. (b) How much kinetic energy in joules is lost during the collision? (c) What percent of the original energy is left?

Solution:

(a) 86.4 N perpendicularly away from the bumper

(b) 0.389 J

(c) 64.0%

Exercise:**Problem:**

During an ice show, a 60.0-kg skater leaps into the air and is caught by an initially stationary 75.0-kg skater. (a) What is their final velocity assuming negligible friction and that the 60.0-kg skater's original horizontal velocity is 4.00 m/s? (b) How much kinetic energy is lost?

Exercise:**Problem: Professional Application**

Using mass and speed data from [\[link\]](#) and assuming that the football player catches the ball with his feet off the ground with both of them moving horizontally, calculate: (a) the final velocity if the ball and player are going in the same direction and (b) the loss of kinetic energy in this case. (c) Repeat parts (a) and (b) for the situation in which the ball and the player are going in opposite directions. Might the loss of kinetic energy be related to how much it hurts to catch the pass?

Solution:

(a) 8.06 m/s

(b) -56.0 J

(c)(i) 7.88 m/s; (ii) -223 J

Exercise:

Problem:

A battleship that is 6.00×10^7 kg and is originally at rest fires a 1100-kg artillery shell horizontally with a velocity of 575 m/s. (a) If the shell is fired straight aft (toward the rear of the ship), there will be negligible friction opposing the ship's recoil. Calculate its recoil velocity. (b) Calculate the increase in internal kinetic energy (that is, for the ship and the shell). This energy is less than the energy released by the gun powder—significant heat transfer occurs.

Exercise:

Problem: Professional Application

Two manned satellites approaching one another, at a relative speed of 0.250 m/s, intending to dock. The first has a mass of 4.00×10^3 kg, and the second a mass of 7.50×10^3 kg. (a) Calculate the final velocity (after docking) by using the frame of reference in which the first satellite was originally at rest. (b) What is the loss of kinetic energy in this inelastic collision? (c) Repeat both parts by using the frame of reference in which the second satellite was originally at rest. Explain why the change in velocity is different in the two frames, whereas the change in kinetic energy is the same in both.

Solution:

(a) 0.163 m/s in the direction of motion of the more massive satellite

(b) 81.6 J

(c) 8.70×10^{-2} m/s in the direction of motion of the less massive satellite, 81.5 J. Because there are no external forces, the velocity of the center of mass of the two-satellite system is unchanged by the collision. The two velocities calculated above are the velocity of the center of mass in each of the two different individual reference frames. The loss in KE is the same in both reference frames because the KE lost to internal forces (heat, friction, etc.) is the same regardless of the coordinate system chosen.

Exercise:

Problem: Professional Application

A 30,000-kg freight car is coasting at 0.850 m/s with negligible friction under a hopper that dumps 110,000 kg of scrap metal into it. (a) What is the final velocity of the loaded freight car? (b) How much kinetic energy is lost?

Exercise:

Problem: Professional Application

Space probes may be separated from their launchers by exploding bolts. (They bolt away from one another.) Suppose a 4800-kg satellite uses this method to separate from the 1500-kg remains of its launcher, and that 5000 J of kinetic energy is supplied to the two parts. What are their subsequent velocities using the frame of reference in which they were at rest before separation?

Solution:

0.704 m/s

−2.25 m/s

Exercise:

Problem:

A 0.0250-kg bullet is accelerated from rest to a speed of 550 m/s in a 3.00-kg rifle. The pain of the rifle's kick is much worse if you hold the gun loosely a few centimeters from your shoulder rather than holding it tightly against your shoulder. (a) Calculate the recoil velocity of the rifle if it is held loosely away from the shoulder. (b) How much kinetic energy does the rifle gain? (c) What is the recoil velocity if the rifle is held tightly against the shoulder, making the effective mass 28.0 kg? (d) How much kinetic energy is transferred to the rifle-shoulder combination? The pain is related to the amount of kinetic energy, which is significantly less in this latter situation. (e) Calculate the momentum of a 110-kg football player running at 8.00 m/s. Compare the player's momentum with the momentum of a hard-thrown 0.410-kg football that has a speed of 25.0 m/s. Discuss its relationship to this problem.

Solution:

(a) 4.58 m/s away from the bullet

(b) 31.5 J

(c) -0.491 m/s

(d) 3.38 J

Exercise:**Problem: Professional Application**

One of the waste products of a nuclear reactor is plutonium-239 (^{239}Pu). This nucleus is radioactive and decays by splitting into a helium-4 nucleus and a uranium-235 nucleus ($^4\text{He} + ^{235}\text{U}$), the latter of which is also radioactive and will itself decay some time later. The energy emitted in the plutonium decay is 8.40×10^{-13} J and is entirely converted to kinetic energy of the helium and uranium nuclei.

The mass of the helium nucleus is 6.68×10^{-27} kg, while that of the uranium is 3.92×10^{-25} kg (note that the ratio of the masses is 4 to 235). (a) Calculate the velocities of the two nuclei, assuming the plutonium nucleus is originally at rest. (b) How much kinetic energy does each nucleus carry away? Note that the data given here are accurate to three digits only.

Exercise:

Problem: Professional Application

The Moon's craters are remnants of meteorite collisions. Suppose a fairly large asteroid that has a mass of 5.00×10^{12} kg (about a kilometer across) strikes the Moon at a speed of 15.0 km/s. (a) At what speed does the Moon recoil after the perfectly inelastic collision (the mass of the Moon is 7.36×10^{22} kg)? (b) How much kinetic energy is lost in the collision? Such an event may have been observed by medieval English monks who reported observing a red glow and subsequent haze about the Moon. (c) In October 2009, NASA crashed a rocket into the Moon, and analyzed the plume produced by the impact. (Significant amounts of water were detected.) Answer part (a) and (b) for this real-life experiment. The mass of the rocket was 2000 kg and its speed upon impact was 9000 km/h. How does the plume produced alter these results?

Solution:

(a) 1.02×10^{-6} m/s

(b) 5.63×10^{20} J (almost all KE lost)

(c) Recoil speed is 6.79×10^{-17} m/s, energy lost is 6.25×10^9 J. The plume will not affect the momentum result because the plume is still part of the Moon system. The plume may affect the kinetic energy result because a significant part of the initial kinetic energy may be transferred to the kinetic energy of the plume particles.

Exercise:

Problem: Professional Application

Two football players collide head-on in midair while trying to catch a thrown football. The first player is 95.0 kg and has an initial velocity of 6.00 m/s, while the second player is 115 kg and has an initial velocity of -3.50 m/s. What is their velocity just after impact if they cling together?

Exercise:**Problem:**

What is the speed of a garbage truck that is 1.20×10^4 kg and is initially moving at 25.0 m/s just after it hits and adheres to a trash can that is 80.0 kg and is initially at rest?

Solution:

24.8 m/s

Exercise:**Problem:**

During a circus act, an elderly performer thrills the crowd by catching a cannon ball shot at him. The cannon ball has a mass of 10.0 kg and the horizontal component of its velocity is 8.00 m/s when the 65.0-kg performer catches it. If the performer is on nearly frictionless roller skates, what is his recoil velocity?

Exercise:**Problem:**

(a) During an ice skating performance, an initially motionless 80.0-kg clown throws a fake barbell away. The clown's ice skates allow her to recoil frictionlessly. If the clown recoils with a velocity of 0.500 m/s and the barbell is thrown with a velocity of 10.0 m/s, what is the mass of the barbell? (b) How much kinetic energy is gained by this maneuver? (c) Where does the kinetic energy come from?

Solution:

(a) 4.00 kg

(b) 210 J

(c) The clown does work to throw the barbell, so the kinetic energy comes from the muscles of the clown. The muscles convert the chemical potential energy of ATP into kinetic energy.

Glossary

inelastic collision

a collision in which internal kinetic energy is not conserved

perfectly inelastic collision

a collision in which the colliding objects stick together

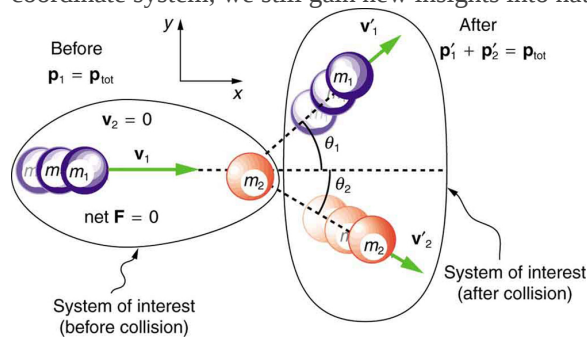
Collisions of Point Masses in Two Dimensions

- Discuss two dimensional collisions as an extension of one dimensional analysis.
- Define point masses.
- Derive an expression for conservation of momentum along x-axis and y-axis.
- Describe elastic collisions of two objects with equal mass.
- Determine the magnitude and direction of the final velocity given initial velocity, and scattering angle.

In the previous two sections, we considered only one-dimensional collisions; during such collisions, the incoming and outgoing velocities are all along the same line. But what about collisions, such as those between billiard balls, in which objects scatter to the side? These are two-dimensional collisions, and we shall see that their study is an extension of the one-dimensional analysis already presented. The approach taken (similar to the approach in discussing two-dimensional kinematics and dynamics) is to choose a convenient coordinate system and resolve the motion into components along perpendicular axes. Resolving the motion yields a pair of one-dimensional problems to be solved simultaneously.

One complication arising in two-dimensional collisions is that the objects might rotate before or after their collision. For example, if two ice skaters hook arms as they pass by one another, they will spin in circles. We will not consider such rotation until later, and so for now we arrange things so that no rotation is possible. To avoid rotation, we consider only the scattering of **point masses**—that is, structureless particles that cannot rotate or spin.

We start by assuming that $\mathbf{F}_{\text{net}} = 0$, so that momentum \mathbf{p} is conserved. The simplest collision is one in which one of the particles is initially at rest. (See [\[link\]](#).) The best choice for a coordinate system is one with an axis parallel to the velocity of the incoming particle, as shown in [\[link\]](#). Because momentum is conserved, the components of momentum along the x - and y -axes (p_x and p_y) will also be conserved, but with the chosen coordinate system, p_y is initially zero and p_x is the momentum of the incoming particle. Both facts simplify the analysis. (Even with the simplifying assumptions of point masses, one particle initially at rest, and a convenient coordinate system, we still gain new insights into nature from the analysis of two-dimensional collisions.)



A two-dimensional collision with the coordinate system chosen so that m_2 is initially at rest and v_1 is parallel to the x -axis. This coordinate system is sometimes called the laboratory coordinate system, because many scattering experiments have a target that is stationary in the laboratory, while particles are scattered from it to determine the particles that make-up the target and how they are bound together. The particles may not be observed directly, but their initial and final velocities are.

Along the x -axis, the equation for conservation of momentum is

Equation:

$$p_{1x} + p_{2x} = p'_{1x} + p'_{2x}.$$

Where the subscripts denote the particles and axes and the primes denote the situation after the collision. In terms of masses and velocities, this equation is

Equation:

$$m_1 v_{1x} + m_2 v_{2x} = m_1 v'_{1x} + m_2 v'_{2x}.$$

But because particle 2 is initially at rest, this equation becomes

Equation:

$$m_1 v_{1x} = m_1 v'_{1x} + m_2 v'_{2x}.$$

The components of the velocities along the x -axis have the form $v \cos \theta$. Because particle 1 initially moves along the x -axis, we find $v_{1x} = v_1$.

Conservation of momentum along the x -axis gives the following equation:

Equation:

$$m_1 v_1 = m_1 v'_1 \cos \theta_1 + m_2 v'_2 \cos \theta_2,$$

where θ_1 and θ_2 are as shown in [\[link\]](#).

Note:

Conservation of Momentum along the x -axis

Equation:

$$m_1 v_1 = m_1 v'_1 \cos \theta_1 + m_2 v'_2 \cos \theta_2$$

Along the y -axis, the equation for conservation of momentum is

Equation:

$$p_{1y} + p_{2y} = p'_{1y} + p'_{2y}$$

or

Equation:

$$m_1 v_{1y} + m_2 v_{2y} = m_1 v'_{1y} + m_2 v'_{2y}.$$

But v_{1y} is zero, because particle 1 initially moves along the x -axis. Because particle 2 is initially at rest, v_{2y} is also zero. The equation for conservation of momentum along the y -axis becomes

Equation:

$$0 = m_1 v'_{1y} + m_2 v'_{2y}.$$

The components of the velocities along the y -axis have the form $v \sin \theta$.

Thus, conservation of momentum along the y -axis gives the following equation:

Equation:

$$0 = m_1 v'_1 \sin \theta_1 + m_2 v'_2 \sin \theta_2.$$

Note:

Conservation of Momentum along the y -axis

Equation:

$$0 = m_1 v'_1 \sin \theta_1 + m_2 v'_2 \sin \theta_2$$

The equations of conservation of momentum along the x -axis and y -axis are very useful in analyzing two-dimensional collisions of particles, where one is originally stationary (a common laboratory situation). But two equations can only be used to find two unknowns, and so other data may be necessary when collision experiments are used to explore nature at the subatomic level.

Example:

Determining the Final Velocity of an Unseen Object from the Scattering of Another Object

Suppose the following experiment is performed. A 0.250-kg object (m_1) is slid on a frictionless surface into a dark room, where it strikes an initially stationary object with mass of 0.400 kg (m_2). The 0.250-kg object emerges from the room at an angle of 45.0° with its incoming direction.

The speed of the 0.250-kg object is originally 2.00 m/s and is 1.50 m/s after the collision. Calculate the magnitude and direction of the velocity (v'_2 and θ_2) of the 0.400-kg object after the collision.

Strategy

Momentum is conserved because the surface is frictionless. The coordinate system shown in [\[link\]](#) is one in which m_2 is originally at rest and the initial velocity is parallel to the x -axis, so that conservation of momentum along the x - and y -axes is applicable.

Everything is known in these equations except v'_2 and θ_2 , which are precisely the quantities we wish to find. We can find two unknowns because we have two independent equations: the equations describing the conservation of momentum in the x - and y -directions.

Solution

Solving $m_1 v_1 = m_1 v'_1 \cos \theta_1 + m_2 v'_2 \cos \theta_2$ for $v'_2 \cos \theta_2$ and $0 = m_1 v'_1 \sin \theta_1 + m_2 v'_2 \sin \theta_2$ for $v'_2 \sin \theta_2$ and taking the ratio yields an equation (in which θ_2 is the only unknown quantity. Applying the identity ($\tan \theta = \frac{\sin \theta}{\cos \theta}$), we obtain:

Equation:

$$\tan \theta_2 = \frac{v'_1 \sin \theta_1}{v'_1 \cos \theta_1 - v_1}.$$

Entering known values into the previous equation gives

Equation:

$$\tan \theta_2 = \frac{(1.50 \text{ m/s})(0.7071)}{(1.50 \text{ m/s})(0.7071) - 2.00 \text{ m/s}} = -1.129.$$

Thus,

Equation:

$$\theta_2 = \tan^{-1}(-1.129) = 311.5^\circ \approx 312^\circ.$$

Angles are defined as positive in the counter clockwise direction, so this angle indicates that m_2 is scattered to the right in [\[link\]](#), as expected (this angle is in the fourth quadrant). Either equation for the x - or y -axis can now be used to solve for v'_2 , but the latter equation is easiest because it has fewer terms.

Equation:

$$v'_2 = -\frac{m_1}{m_2} v_1 \frac{\sin \theta_1}{\sin \theta_2}$$

Entering known values into this equation gives

Equation:

$$v'_2 = -\left(\frac{0.250 \text{ kg}}{0.400 \text{ kg}}\right)(1.50 \text{ m/s})\left(\frac{0.7071}{-0.7485}\right).$$

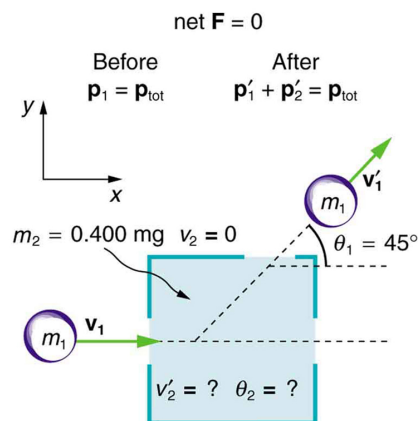
Thus,

Equation:

$$v'_2 = 0.886 \text{ m/s}.$$

Discussion

It is instructive to calculate the internal kinetic energy of this two-object system before and after the collision. (This calculation is left as an end-of-chapter problem.) If you do this calculation, you will find that the internal kinetic energy is less after the collision, and so the collision is inelastic. This type of result makes a physicist want to explore the system further.



A collision taking place in a dark room is explored in [\[link\]](#).

The incoming object m_1 is scattered by an initially stationary object. Only the stationary object's mass m_2 is known. By measuring the angle and speed at which m_1 emerges from the room, it is possible to calculate the magnitude and

direction of the initially
stationary object's velocity after
the collision.

Elastic Collisions of Two Objects with Equal Mass

Some interesting situations arise when the two colliding objects have equal mass and the collision is elastic. This situation is nearly the case with colliding billiard balls, and precisely the case with some subatomic particle collisions. We can thus get a mental image of a collision of subatomic particles by thinking about billiards (or pool). (Refer to [\[link\]](#) for masses and angles.) First, an elastic collision conserves internal kinetic energy. Again, let us assume object 2 (m_2) is initially at rest. Then, the internal kinetic energy before and after the collision of two objects that have equal masses is

Equation:

$$\frac{1}{2}mv_1^2 = \frac{1}{2}mv_1'^2 + \frac{1}{2}mv_2'^2.$$

Because the masses are equal, $m_1 = m_2 = m$. Algebraic manipulation (left to the reader) of conservation of momentum in the x - and y -directions can show that

Equation:

$$\frac{1}{2}mv_1^2 = \frac{1}{2}mv_1'^2 + \frac{1}{2}mv_2'^2 + mv_1v_2' \cos(\theta_1 - \theta_2).$$

(Remember that θ_2 is negative here.) The two preceding equations can both be true only if

Equation:

$$mv_1v_2' \cos(\theta_1 - \theta_2) = 0.$$

There are three ways that this term can be zero. They are

- $v_1 = 0$: head-on collision; incoming ball stops
- $v_2' = 0$: no collision; incoming ball continues unaffected
- $\cos(\theta_1 - \theta_2) = 0$: angle of separation ($\theta_1 - \theta_2$) is 90° after the collision

All three of these ways are familiar occurrences in billiards and pool, although most of us try to avoid the second. If you play enough pool, you will notice that the angle between the balls is very close to 90° after the collision, although it will vary from this value if a great deal of spin is placed on the ball. (Large spin carries in extra energy and a quantity called *angular momentum*, which must also be conserved.) The assumption that the scattering of billiard balls is elastic is reasonable based on the correctness of the three results it produces. This assumption also implies that, to a good approximation, momentum is conserved for the two-ball system in billiards and pool. The problems below explore these and other characteristics of two-dimensional collisions.

Note:

Connections to Nuclear and Particle Physics

Two-dimensional collision experiments have revealed much of what we know about subatomic particles, as we shall see in [Medical Applications of Nuclear Physics](#) and [Particle Physics](#). Ernest Rutherford, for example, discovered the nature of the atomic nucleus from such experiments.

Section Summary

- The approach to two-dimensional collisions is to choose a convenient coordinate system and break the motion into components along perpendicular axes. Choose a coordinate system with the x -axis parallel to the velocity of the incoming particle.
 - Two-dimensional collisions of point masses where mass 2 is initially at rest conserve momentum along the initial direction of mass 1 (the x -axis), stated by $m_1 v_1 = m_1 v'_1 \cos \theta_1 + m_2 v'_2 \cos \theta_2$ and along the direction perpendicular to the initial direction (the y -axis) stated by $0 = m_1 v'_{1y} + m_2 v'_{2y}$.
 - The internal kinetic before and after the collision of two objects that have equal masses is
- Equation:**

$$\frac{1}{2} m v_1^2 = \frac{1}{2} m v'^2_1 + \frac{1}{2} m v'^2_2 + m v_1 v'_2 \cos(\theta_1 - \theta_2).$$

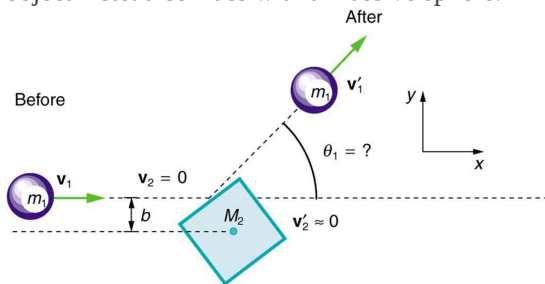
- Point masses are structureless particles that cannot spin.

Conceptual Questions

Exercise:

Problem:

[\[link\]](#) shows a cube at rest and a small object heading toward it. (a) Describe the directions (angle θ_1) at which the small object can emerge after colliding elastically with the cube. How does θ_1 depend on b , the so-called impact parameter? Ignore any effects that might be due to rotation after the collision, and assume that the cube is much more massive than the small object. (b) Answer the same questions if the small object instead collides with a massive sphere.



A small object approaches a collision with a much more massive cube, after which its velocity has the direction θ_1 . The angles at which the small object can be scattered are determined by the shape of the object it strikes and the impact parameter b .

Problems & Exercises

Exercise:

Problem:

Two identical pucks collide on an air hockey table. One puck was originally at rest. (a) If the incoming puck has a speed of 6.00 m/s and scatters to an angle of 30.0° , what is the velocity (magnitude and direction) of the second puck? (You may use the result that $\theta_1 - \theta_2 = 90^\circ$ for elastic collisions of objects that have identical masses.) (b) Confirm that the collision is elastic.

Solution:

(a) 3.00 m/s, 60° below x -axis

(b) Find speed of first puck after collision:

$$0 = mv_1' \sin 30^\circ - mv_2' \sin 60^\circ \Rightarrow v_1' = v_2' \frac{\sin 60^\circ}{\sin 30^\circ} = 5.196 \text{ m/s}$$

Verify that ratio of initial to final KE equals one:
$$\left. \begin{aligned} \text{KE} &= \frac{1}{2}mv_1^2 = 18m \text{ J} \\ \text{KE} &= \frac{1}{2}mv_1'^2 + \frac{1}{2}mv_2'^2 = 18m \text{ J} \end{aligned} \right\} \frac{\text{KE}}{\text{KE}} = 1.00$$

Exercise:**Problem:**

Confirm that the results of the example [\[link\]](#) do conserve momentum in both the x - and y -directions.

Exercise:**Problem:**

A 3000-kg cannon is mounted so that it can recoil only in the horizontal direction. (a) Calculate its recoil velocity when it fires a 15.0-kg shell at 480 m/s at an angle of 20.0° above the horizontal. (b) What is the kinetic energy of the cannon? This energy is dissipated as heat transfer in shock absorbers that stop its recoil. (c) What happens to the vertical component of momentum that is imparted to the cannon when it is fired?

Solution:

(a) -2.26 m/s

(b) $7.63 \times 10^3 \text{ J}$

(c) The ground will exert a normal force to oppose recoil of the cannon in the vertical direction. The momentum in the vertical direction is transferred to the earth. The energy is transferred into the ground, making a dent where the cannon is. After long barrages, cannon have erratic aim because the ground is full of divots.

Exercise:**Problem: Professional Application**

A 5.50-kg bowling ball moving at 9.00 m/s collides with a 0.850-kg bowling pin, which is scattered at an angle of 85.0° to the initial direction of the bowling ball and with a speed of 15.0 m/s. (a) Calculate the final velocity (magnitude and direction) of the bowling ball. (b) Is the collision elastic? (c) Linear kinetic energy is greater after the collision. Discuss how spin on the ball might be converted to linear kinetic energy in the collision.

Exercise:**Problem: Professional Application**

Ernest Rutherford (the first New Zealander to be awarded the Nobel Prize in Chemistry) demonstrated that nuclei were very small and dense by scattering helium-4 nuclei (${}^4\text{He}$) from gold-197 nuclei (${}^{197}\text{Au}$). The energy of the incoming helium nucleus was $8.00 \times 10^{-13} \text{ J}$, and the masses of the helium and gold nuclei were $6.68 \times 10^{-27} \text{ kg}$ and $3.29 \times 10^{-25} \text{ kg}$, respectively (note that their mass ratio is 4 to 197). (a) If a helium nucleus scatters to an angle of 120° during an elastic collision with a gold nucleus, calculate the helium nucleus's final speed and the final velocity (magnitude and direction) of the gold nucleus. (b) What is the final kinetic energy of the helium nucleus?

Solution:

(a) $5.36 \times 10^5 \text{ m/s}$ at -29.5°

(b) $7.52 \times 10^{-13} \text{ J}$

Exercise:

Problem: Professional Application

Two cars collide at an icy intersection and stick together afterward. The first car has a mass of 1200 kg and is approaching at 8.00 m/s due south. The second car has a mass of 850 kg and is approaching at 17.0 m/s due west. (a) Calculate the final velocity (magnitude and direction) of the cars. (b) How much kinetic energy is lost in the collision? (This energy goes into deformation of the cars.) Note that because both cars have an initial velocity, you cannot use the equations for conservation of momentum along the x -axis and y -axis; instead, you must look for other simplifying aspects.

Exercise:

Problem:

Starting with equations $m_1 v_1 = m_1 v'_1 \cos \theta_1 + m_2 v'_2 \cos \theta_2$ and $0 = m_1 v'_1 \sin \theta_1 + m_2 v'_2 \sin \theta_2$ for conservation of momentum in the x - and y -directions and assuming that one object is originally stationary, prove that for an elastic collision of two objects of equal masses,

Equation:

$$\frac{1}{2} m v_1^2 = \frac{1}{2} m v'^2_1 + \frac{1}{2} m v'^2_2 + m v'_1 v'_2 \cos (\theta_1 - \theta_2)$$

as discussed in the text.

Solution:

We are given that $m_1 = m_2 \equiv m$. The given equations then become:

Equation:

$$v_1 = v'_1 \cos \theta_1 + v'_2 \cos \theta_2$$

and

Equation:

$$0 = v'_1 \sin \theta_1 + v'_2 \sin \theta_2.$$

Square each equation to get

Equation:

$$\begin{aligned}
 v_1^2 &= v_1'^2 \cos^2 \theta_1 + v_2'^2 \cos^2 \theta_2 + 2v_1'v_2' \cos \theta_1 \cos \theta_2 \\
 0 &= v_1'^2 \sin^2 \theta_1 + v_2'^2 \sin^2 \theta_2 + 2v_1'v_2' \sin \theta_1 \sin \theta_2.
 \end{aligned}$$

Add these two equations and simplify:

Equation:

$$\begin{aligned}
 v_1^2 &= v_1'^2 + v_2'^2 + 2v_1'v_2'(\cos \theta_1 \cos \theta_2 + \sin \theta_1 \sin \theta_2) \\
 &= v_1'^2 + v_2'^2 + 2v_1'v_2' \left[\frac{1}{2} \cos (\theta_1 - \theta_2) + \frac{1}{2} \cos (\theta_1 + \theta_2) + \frac{1}{2} \cos (\theta_1 - \theta_2) - \frac{1}{2} \cos (\theta_1 + \theta_2) \right] \\
 &= v_1'^2 + v_2'^2 + 2v_1'v_2' \cos (\theta_1 - \theta_2).
 \end{aligned}$$

Multiply the entire equation by $\frac{1}{2}m$ to recover the kinetic energy:

Equation:

$$\frac{1}{2}mv_1^2 = \frac{1}{2}mv_1'^2 + \frac{1}{2}mv_2'^2 + mv_1'v_2' \cos(\theta_1 - \theta_2)$$

Exercise:

Problem: Integrated Concepts

A 90.0-kg ice hockey player hits a 0.150-kg puck, giving the puck a velocity of 45.0 m/s. If both are initially at rest and if the ice is frictionless, how far does the player recoil in the time it takes the puck to reach the goal 15.0 m away?

Glossary

point masses

structureless particles with no rotation or spin

Introduction to Rocket Propulsion

- State Newton's third law of motion.
- Explain the principle involved in propulsion of rockets and jet engines.
- Derive an expression for the acceleration of the rocket and discuss the factors that affect the acceleration.
- Describe the function of a space shuttle.

Rockets range in size from fireworks so small that ordinary people use them to immense Saturn Vs that once propelled massive payloads toward the Moon. The propulsion of all rockets, jet engines, deflating balloons, and even squids and octopuses is explained by the same physical principle—Newton's third law of motion. Matter is forcefully ejected from a system, producing an equal and opposite reaction on what remains. Another common example is the recoil of a gun. The gun exerts a force on a bullet to accelerate it and consequently experiences an equal and opposite force, causing the gun's recoil or kick.

Note:

Making Connections: Take-Home Experiment—Propulsion of a Balloon

Hold a balloon and fill it with air. Then, let the balloon go. In which direction does the air come out of the balloon and in which direction does the balloon get propelled? If you fill the balloon with water and then let the balloon go, does the balloon's direction change? Explain your answer.

[\[link\]](#) shows a rocket accelerating straight up. In part (a), the rocket has a mass m and a velocity v relative to Earth, and hence a momentum mv . In part (b), a time Δt has elapsed in which the rocket has ejected a mass Δm of hot gas at a velocity v_e relative to the rocket. The remainder of the mass ($m - \Delta m$) now has a greater velocity ($v + \Delta v$). The momentum of the entire system (rocket plus expelled gas) has actually decreased because the force of gravity has acted for a time Δt , producing a negative impulse $\Delta p = -mg\Delta t$. (Remember that impulse is the net external force on a system multiplied by the time it acts, and it equals the change in momentum

of the system.) So, the center of mass of the system is in free fall but, by rapidly expelling mass, part of the system can accelerate upward. It is a commonly held misconception that the rocket exhaust pushes on the ground. If we consider thrust; that is, the force exerted on the rocket by the exhaust gases, then a rocket's thrust is greater in outer space than in the atmosphere or on the launch pad. In fact, gases are easier to expel into a vacuum.

By calculating the change in momentum for the entire system over Δt , and equating this change to the impulse, the following expression can be shown to be a good approximation for the acceleration of the rocket.

Equation:

$$a = \frac{v_e}{m} \frac{\Delta m}{\Delta t} - g$$

“The rocket” is that part of the system remaining after the gas is ejected, and g is the acceleration due to gravity.

Note:

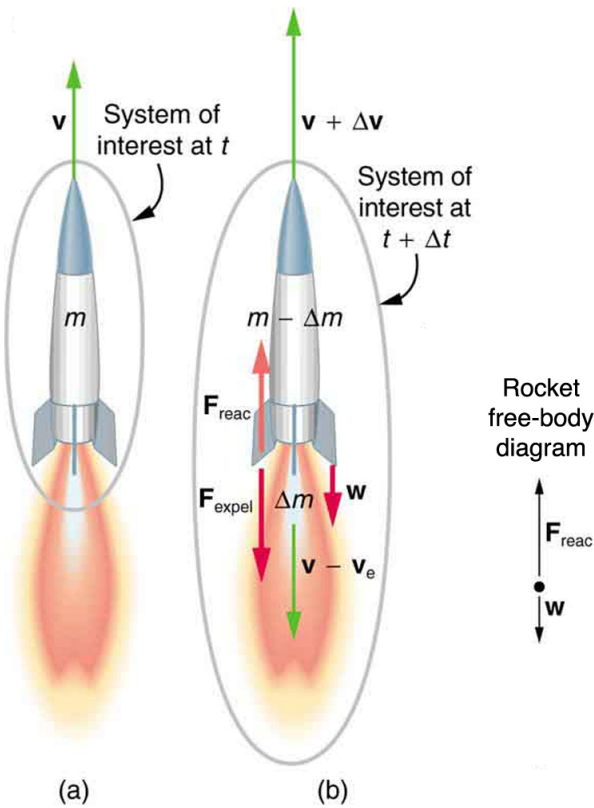
Acceleration of a Rocket

Acceleration of a rocket is

Equation:

$$a = \frac{v_e}{m} \frac{\Delta m}{\Delta t} - g,$$

where a is the acceleration of the rocket, v_e is the exhaust velocity, m is the mass of the rocket, Δm is the mass of the ejected gas, and Δt is the time in which the gas is ejected.



(a) This rocket has a mass m and an upward velocity v . The net external force on the system is $-mg$, if air resistance is neglected. (b) A time Δt later the system has two main parts, the ejected gas and the remainder of the rocket. The reaction force on the rocket is what overcomes the gravitational force and accelerates it upward.

A rocket's acceleration depends on three major factors, consistent with the equation for acceleration of a rocket . First, the greater the exhaust velocity of the gases relative to the rocket, v_e , the greater the acceleration is. The

practical limit for v_e is about 2.5×10^3 m/s for conventional (non-nuclear) hot-gas propulsion systems. The second factor is the rate at which mass is ejected from the rocket. This is the factor $\Delta m / \Delta t$ in the equation. The quantity $(\Delta m / \Delta t)v_e$, with units of newtons, is called "thrust." The faster the rocket burns its fuel, the greater its thrust, and the greater its acceleration. The third factor is the mass m of the rocket. The smaller the mass is (all other factors being the same), the greater the acceleration. The rocket mass m decreases dramatically during flight because most of the rocket is fuel to begin with, so that acceleration increases continuously, reaching a maximum just before the fuel is exhausted.

Note:**Factors Affecting a Rocket's Acceleration**

- The greater the exhaust velocity v_e of the gases relative to the rocket, the greater the acceleration.
- The faster the rocket burns its fuel, the greater its acceleration.
- The smaller the rocket's mass (all other factors being the same), the greater the acceleration.

Example:**Calculating Acceleration: Initial Acceleration of a Moon Launch**

A Saturn V's mass at liftoff was 2.80×10^6 kg, its fuel-burn rate was 1.40×10^4 kg/s, and the exhaust velocity was 2.40×10^3 m/s. Calculate its initial acceleration.

Strategy

This problem is a straightforward application of the expression for acceleration because a is the unknown and all of the terms on the right side of the equation are given.

Solution

Substituting the given values into the equation for acceleration yields

Equation:

$$\begin{aligned}
 a &= \frac{v_e}{m} \frac{\Delta m}{\Delta t} - g \\
 &= \frac{2.40 \times 10^3 \text{ m/s}}{2.80 \times 10^6 \text{ kg}} (1.40 \times 10^4 \text{ kg/s}) - 9.80 \text{ m/s}^2 \\
 &= 2.20 \text{ m/s}^2.
 \end{aligned}$$

Discussion

This value is fairly small, even for an initial acceleration. The acceleration does increase steadily as the rocket burns fuel, because m decreases while v_e and $\frac{\Delta m}{\Delta t}$ remain constant. Knowing this acceleration and the mass of the rocket, you can show that the thrust of the engines was $3.36 \times 10^7 \text{ N}$.

To achieve the high speeds needed to hop continents, obtain orbit, or escape Earth's gravity altogether, the mass of the rocket other than fuel must be as small as possible. It can be shown that, in the absence of air resistance and neglecting gravity, the final velocity of a one-stage rocket initially at rest is

Equation:

$$v = v_e \ln \frac{m_0}{m_r},$$

where $\ln(m_0/m_r)$ is the natural logarithm of the ratio of the initial mass of the rocket (m_0) to what is left (m_r) after all of the fuel is exhausted. (Note that v is actually the change in velocity, so the equation can be used for any segment of the flight. If we start from rest, the change in velocity equals the final velocity.) For example, let us calculate the mass ratio needed to escape Earth's gravity starting from rest, given that the escape velocity from Earth is about $11.2 \times 10^3 \text{ m/s}$, and assuming an exhaust velocity $v_e = 2.5 \times 10^3 \text{ m/s}$.

Equation:

$$\ln \frac{m_0}{m_r} = \frac{v}{v_e} = \frac{11.2 \times 10^3 \text{ m/s}}{2.5 \times 10^3 \text{ m/s}} = 4.48$$

Solving for m_0/m_r gives

Equation:

$$\frac{m_0}{m_r} = e^{4.48} = 88.$$

Thus, the mass of the rocket is

Equation:

$$m_r = \frac{m_0}{88}.$$

This result means that only 1/88 of the mass is left when the fuel is burnt, and 87/88 of the initial mass was fuel. Expressed as percentages, 98.9% of the rocket is fuel, while payload, engines, fuel tanks, and other components make up only 1.10%. Taking air resistance and gravitational force into account, the mass m_r remaining can only be about $m_0/180$. It is difficult to build a rocket in which the fuel has a mass 180 times everything else. The solution is multistage rockets. Each stage only needs to achieve part of the final velocity and is discarded after it burns its fuel. The result is that each successive stage can have smaller engines and more payload relative to its fuel. Once out of the atmosphere, the ratio of payload to fuel becomes more favorable, too.

The space shuttle was an attempt at an economical vehicle with some reusable parts, such as the solid fuel boosters and the craft itself. (See [\[link\]](#)) The shuttle's need to be operated by humans, however, made it at least as costly for launching satellites as expendable, unmanned rockets. Ideally, the shuttle would only have been used when human activities were required for the success of a mission, such as the repair of the Hubble space telescope. Rockets with satellites can also be launched from airplanes. Using airplanes has the double advantage that the initial velocity is significantly above zero and a rocket can avoid most of the atmosphere's resistance.



The space shuttle had a number of reusable parts.

Solid fuel boosters on either side were recovered and refueled after each flight, and the entire orbiter returned to Earth for use in subsequent flights. The large liquid fuel tank was expended. The space shuttle was a complex assemblage of technologies, employing both solid and liquid fuel and pioneering ceramic tiles as reentry heat shields. As a result, it permitted multiple launches as opposed to

single-use rockets.
(credit: NASA)

Note:

PhET Explorations: Lunar Lander

Can you avoid the boulder field and land safely, just before your fuel runs out, as Neil Armstrong did in 1969? Our version of this classic video game accurately simulates the real motion of the lunar lander with the correct mass, thrust, fuel consumption rate, and lunar gravity. The real lunar lander is very hard to control.

https://phet.colorado.edu/sims/lunar-lander/lunar-lander_en.html

Section Summary

- Newton's third law of motion states that to every action, there is an equal and opposite reaction.
- Acceleration of a rocket is $a = \frac{v_e}{m} \frac{\Delta m}{\Delta t} - g$.
- A rocket's acceleration depends on three main factors. They are
 1. The greater the exhaust velocity of the gases, the greater the acceleration.
 2. The faster the rocket burns its fuel, the greater its acceleration.
 3. The smaller the rocket's mass, the greater the acceleration.

Conceptual Questions

Exercise:

Problem: Professional Application

Suppose a fireworks shell explodes, breaking into three large pieces for which air resistance is negligible. How is the motion of the center

of mass affected by the explosion? How would it be affected if the pieces experienced significantly more air resistance than the intact shell?

Exercise:

Problem: Professional Application

During a visit to the International Space Station, an astronaut was positioned motionless in the center of the station, out of reach of any solid object on which he could exert a force. Suggest a method by which he could move himself away from this position, and explain the physics involved.

Exercise:

Problem: Professional Application

It is possible for the velocity of a rocket to be greater than the exhaust velocity of the gases it ejects. When that is the case, the gas velocity and gas momentum are in the same direction as that of the rocket. How is the rocket still able to obtain thrust by ejecting the gases?

Problems & Exercises

Exercise:

Problem: Professional Application

Antiballistic missiles (ABMs) are designed to have very large accelerations so that they may intercept fast-moving incoming missiles in the short time available. What is the takeoff acceleration of a 10,000-kg ABM that expels 196 kg of gas per second at an exhaust velocity of $2.50 \times 10^3 \text{ m/s}$?

Solution:

$$39.2 \text{ m/s}^2$$

Exercise:

Problem: Professional Application

What is the acceleration of a 5000-kg rocket taking off from the Moon, where the acceleration due to gravity is only 1.6 m/s^2 , if the rocket expels 8.00 kg of gas per second at an exhaust velocity of $2.20 \times 10^3 \text{ m/s}$?

Exercise:

Problem: Professional Application

Calculate the increase in velocity of a 4000-kg space probe that expels 3500 kg of its mass at an exhaust velocity of $2.00 \times 10^3 \text{ m/s}$. You may assume the gravitational force is negligible at the probe's location.

Solution:

$$4.16 \times 10^3 \text{ m/s}$$

Exercise:

Problem: Professional Application

Ion-propulsion rockets have been proposed for use in space. They employ atomic ionization techniques and nuclear energy sources to produce extremely high exhaust velocities, perhaps as great as $8.00 \times 10^6 \text{ m/s}$. These techniques allow a much more favorable payload-to-fuel ratio. To illustrate this fact: (a) Calculate the increase in velocity of a 20,000-kg space probe that expels only 40.0-kg of its mass at the given exhaust velocity. (b) These engines are usually designed to produce a very small thrust for a very long time—the type of engine that might be useful on a trip to the outer planets, for example. Calculate the acceleration of such an engine if it expels

$4.50 \times 10^{-6} \text{ kg/s}$ at the given velocity, assuming the acceleration due to gravity is negligible.

Exercise:

Problem: Derive the equation for the vertical acceleration of a rocket.

Solution:

The force needed to give a small mass Δm an acceleration $a_{\Delta m}$ is $F = \Delta m a_{\Delta m}$. To accelerate this mass in the small time interval Δt at a speed v_e requires $v_e = a_{\Delta m} \Delta t$, so $F = v_e \frac{\Delta m}{\Delta t}$. By Newton's third law, this force is equal in magnitude to the thrust force acting on the rocket, so $F_{\text{thrust}} = v_e \frac{\Delta m}{\Delta t}$, where all quantities are positive. Applying Newton's second law to the rocket gives $F_{\text{thrust}} - mg = ma \Rightarrow a = \frac{v_e}{m} \frac{\Delta m}{\Delta t} - g$, where m is the mass of the rocket and unburnt fuel.

Exercise:

Problem: Professional Application

(a) Calculate the maximum rate at which a rocket can expel gases if its acceleration cannot exceed seven times that of gravity. The mass of the rocket just as it runs out of fuel is 75,000-kg, and its exhaust velocity is $2.40 \times 10^3 \text{ m/s}$. Assume that the acceleration of gravity is the same as on Earth's surface (9.80 m/s^2). (b) Why might it be necessary to limit the acceleration of a rocket?

Exercise:

Problem:

Given the following data for a fire extinguisher-toy wagon rocket experiment, calculate the average exhaust velocity of the gases expelled from the extinguisher. Starting from rest, the final velocity is 10.0 m/s. The total mass is initially 75.0 kg and is 70.0 kg after the extinguisher is fired.

Exercise:**Problem:**

How much of a single-stage rocket that is 100,000 kg can be anything but fuel if the rocket is to have a final speed of 8.00 km/s, given that it expels gases at an exhaust velocity of 2.20×10^3 m/s?

Solution:

$$2.63 \times 10^3 \text{ kg}$$

Exercise:**Problem: Professional Application**

(a) A 5.00-kg squid initially at rest ejects 0.250-kg of fluid with a velocity of 10.0 m/s. What is the recoil velocity of the squid if the ejection is done in 0.100 s and there is a 5.00-N frictional force opposing the squid's movement. (b) How much energy is lost to work done against friction?

Solution:

(a) 0.421 m/s away from the ejected fluid.

(b) 0.237 J.

Exercise:**Problem: Unreasonable Results**

Squids have been reported to jump from the ocean and travel 30.0 m (measured horizontally) before re-entering the water. (a) Calculate the initial speed of the squid if it leaves the water at an angle of 20.0° , assuming negligible lift from the air and negligible air resistance. (b) The squid propels itself by squirting water. What fraction of its mass would it have to eject in order to achieve the speed found in the previous part? The water is ejected at 12.0 m/s; gravitational force and friction are neglected. (c) What is unreasonable about the results? (d) Which premise is unreasonable, or which premises are inconsistent?

Exercise:

Problem: Construct Your Own Problem

Consider an astronaut in deep space cut free from her space ship and needing to get back to it. The astronaut has a few packages that she can throw away to move herself toward the ship. Construct a problem in which you calculate the time it takes her to get back by throwing all the packages at one time compared to throwing them one at a time. Among the things to be considered are the masses involved, the force she can exert on the packages through some distance, and the distance to the ship.

Exercise:

Problem: Construct Your Own Problem

Consider an artillery projectile striking armor plating. Construct a problem in which you find the force exerted by the projectile on the plate. Among the things to be considered are the mass and speed of the projectile and the distance over which its speed is reduced. Your instructor may also wish for you to consider the relative merits of depleted uranium versus lead projectiles based on the greater density of uranium.

Introduction to Temperature, Kinetic Theory, and the Gas Laws

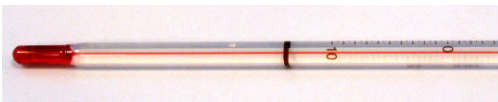
class="introduction"

The welder's
gloves and
helmet
protect him
from the
electric arc
that transfers
enough
thermal
energy to
melt the rod,
spray sparks,
and burn the
retina of an
unprotected
eye. The
thermal
energy can
be felt on
exposed skin
a few meters
away, and its
light can be
seen for
kilometers.
(credit:
Kevin S.
O'Brien/U.S
. Navy)



Heat is something familiar to each of us. We feel the warmth of the summer Sun, the chill of a clear summer night, the heat of coffee after a winter stroll, and the cooling effect of our sweat. Heat transfer is maintained by temperature differences. Manifestations of **heat transfer**—the movement of heat energy from one place or material to another—are apparent throughout the universe. Heat from beneath Earth's surface is brought to the surface in flows of incandescent lava. The Sun warms Earth's surface and is the source of much of the energy we find on it. Rising levels of atmospheric carbon dioxide threaten to trap more of the Sun's energy, perhaps fundamentally altering the ecosphere. In space, supernovas explode, briefly radiating more heat than an entire galaxy does.

What is heat? How do we define it? How is it related to temperature? What are heat's effects? How is it related to other forms of energy and to work? We will find that, in spite of the richness of the phenomena, there is a small set of underlying physical principles that unite the subjects and tie them to other fields.



In a typical thermometer like this one, the alcohol, with a red dye, expands

more rapidly than the glass containing it. When the thermometer's temperature increases, the liquid from the bulb is forced into the narrow tube, producing a large change in the length of the column for a small change in temperature.

(credit: Chemical Engineer, Wikimedia Commons)

Temperature

- Define temperature.
- Convert temperatures between the Celsius, Fahrenheit, and Kelvin scales.
- Define thermal equilibrium.
- State the zeroth law of thermodynamics.

The concept of temperature has evolved from the common concepts of hot and cold. Human perception of what feels hot or cold is a relative one. For example, if you place one hand in hot water and the other in cold water, and then place both hands in tepid water, the tepid water will feel cool to the hand that was in hot water, and warm to the one that was in cold water. The scientific definition of temperature is less ambiguous than your senses of hot and cold. **Temperature** is operationally defined to be what we measure with a thermometer. (Many physical quantities are defined solely in terms of how they are measured. We shall see later how temperature is related to the kinetic energies of atoms and molecules, a more physical explanation.) Two accurate thermometers, one placed in hot water and the other in cold water, will show the hot water to have a higher temperature. If they are then placed in the tepid water, both will give identical readings (within measurement uncertainties). In this section, we discuss temperature, its measurement by thermometers, and its relationship to thermal equilibrium. Again, temperature is the quantity measured by a thermometer.

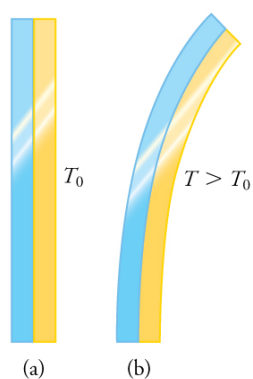
Note:

Misconception Alert: Human Perception vs. Reality

On a cold winter morning, the wood on a porch feels warmer than the metal of your bike. The wood and bicycle are in thermal equilibrium with the outside air, and are thus the same temperature. They *feel* different because of the difference in the way that they conduct heat away from your skin. The metal conducts heat away from your body faster than the wood does (see more about conductivity in [Conduction](#)). This is just one example demonstrating that the human sense of hot and cold is not determined by temperature alone.

Another factor that affects our perception of temperature is humidity. Most people feel much hotter on hot, humid days than on hot, dry days. This is because on humid days, sweat does not evaporate from the skin as efficiently as it does on dry days. It is the evaporation of sweat (or water from a sprinkler or pool) that cools us off.

Any physical property that depends on temperature, and whose response to temperature is reproducible, can be used as the basis of a thermometer. Because many physical properties depend on temperature, the variety of thermometers is remarkable. For example, volume increases with temperature for most substances. This property is the basis for the common alcohol thermometer, the old mercury thermometer, and the bimetallic strip ([link](#)). Other properties used to measure temperature include electrical resistance and color, as shown in [link](#), and the emission of infrared radiation, as shown in [link](#).



The curvature of a bimetallic strip depends on

temperature.
(a) The strip is straight at the starting temperature, where its two components have the same length.

(b) At a higher temperature, this strip bends to the right, because the metal on the left has expanded more than the metal on the right.



Each of the six squares on this plastic (liquid crystal)

thermometer contains a film of a different heat-sensitive liquid crystal material. Below 95°F, all six squares are black.

When the plastic thermometer is exposed to temperature that increases to 95°F, the first liquid crystal square changes color. When the temperature increases above 96.8°F the second liquid crystal square also changes color, and so forth. (credit: Arkrishna, Wikimedia Commons)



Fireman Jason Ormand uses a pyrometer to check the temperature of an aircraft carrier's ventilation system. Infrared radiation (whose emission varies with temperature)

from the vent is measured and a temperature readout is quickly produced.

Infrared measurements are also frequently used as a measure of body temperature.

These modern thermometers, placed in the ear canal, are more accurate than alcohol thermometers placed under the tongue or in the armpit. (credit: Lamel J. Hinton/U.S. Navy)

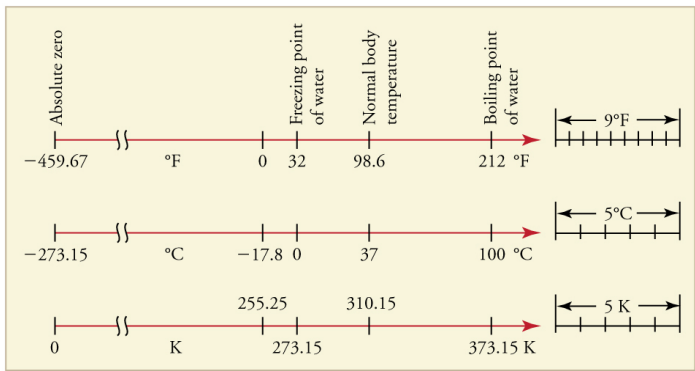
Temperature Scales

Thermometers are used to measure temperature according to well-defined scales of measurement, which use pre-defined reference points to help compare quantities. The three most common temperature scales are the Fahrenheit, Celsius, and Kelvin scales. A temperature scale can be created by identifying two easily reproducible temperatures. The freezing and boiling temperatures of water at standard atmospheric pressure are commonly used.

The **Celsius** scale (which replaced the slightly different *centigrade* scale) has the freezing point of water at 0°C and the boiling point at 100°C . Its unit is the **degree Celsius** ($^{\circ}\text{C}$). On the **Fahrenheit** scale (still the most frequently used in the United States), the freezing point of water is at 32°F and the boiling point is at 212°F . The unit of temperature on this scale is the **degree Fahrenheit** ($^{\circ}\text{F}$). Note that a temperature difference of one degree Celsius is greater than a temperature difference of one degree Fahrenheit. Only 100 Celsius degrees

span the same range as 180 Fahrenheit degrees, thus one degree on the Celsius scale is 1.8 times larger than one degree on the Fahrenheit scale $180/100 = 9/5$.

The **Kelvin** scale is the temperature scale that is commonly used in science. It is an *absolute temperature* scale defined to have 0 K at the lowest possible temperature, called **absolute zero**. The official temperature unit on this scale is the *kelvin*, which is abbreviated K, and is not accompanied by a degree sign. The freezing and boiling points of water are 273.15 K and 373.15 K, respectively. Thus, the magnitude of temperature differences is the same in units of kelvins and degrees Celsius. Unlike other temperature scales, the Kelvin scale is an absolute scale. It is used extensively in scientific work because a number of physical quantities, such as the volume of an ideal gas, are directly related to absolute temperature. The kelvin is the SI unit used in scientific work.



Relationships between the Fahrenheit, Celsius, and Kelvin temperature scales, rounded to the nearest degree. The relative sizes of the scales are also shown.

The relationships between the three common temperature scales is shown in [\[link\]](#). Temperatures on these scales can be converted using the equations in [\[link\]](#).

To convert from ...	Use this equation ...	Also written as ...
---------------------	-----------------------	---------------------

To convert from ...	Use this equation ...	Also written as ...
Celsius to Fahrenheit	$T(^{\circ}\text{F}) = \frac{9}{5}T(^{\circ}\text{C}) + 32$	$T_{\text{F}} = \frac{9}{5}T_{\text{C}} + 32$
Fahrenheit to Celsius	$T(^{\circ}\text{C}) = \frac{5}{9}(T(^{\circ}\text{F}) - 32)$	$T_{\text{C}} = \frac{5}{9}(T_{\text{F}} - 32)$
Celsius to Kelvin	$T(\text{K}) = T(^{\circ}\text{C}) + 273.15$	$T_{\text{K}} = T_{\text{C}} + 273.15$
Kelvin to Celsius	$T(^{\circ}\text{C}) = T(\text{K}) - 273.15$	$T_{\text{C}} = T_{\text{K}} - 273.15$
Fahrenheit to Kelvin	$T(\text{K}) = \frac{5}{9}(T(^{\circ}\text{F}) - 32) + 273.15$	$T_{\text{K}} = \frac{5}{9}(T_{\text{F}} - 32) + 273.15$
Kelvin to Fahrenheit	$T(^{\circ}\text{F}) = \frac{9}{5}(T(\text{K}) - 273.15) + 32$	$T_{\text{F}} = \frac{9}{5}(T_{\text{K}} - 273.15) + 32$

Temperature Conversions

Notice that the conversions between Fahrenheit and Kelvin look quite complicated. In fact, they are simple combinations of the conversions between Fahrenheit and Celsius, and the conversions between Celsius and Kelvin.

Example:

Converting between Temperature Scales: Room Temperature

“Room temperature” is generally defined to be 25°C . (a) What is room temperature in $^{\circ}\text{F}$?

(b) What is it in K?

Strategy

To answer these questions, all we need to do is choose the correct conversion equations and plug in the known values.

Solution for (a)

1. Choose the right equation. To convert from °C to °F, use the equation

Equation:

$$T_{\text{°F}} = \frac{9}{5}T_{\text{°C}} + 32.$$

2. Plug the known value into the equation and solve:

Equation:

$$T_{\text{°F}} = \frac{9}{5}25^{\circ}\text{C} + 32 = 77^{\circ}\text{F}.$$

Solution for (b)

1. Choose the right equation. To convert from °C to K, use the equation

Equation:

$$T_{\text{K}} = T_{\text{°C}} + 273.15.$$

2. Plug the known value into the equation and solve:

Equation:

$$T_{\text{K}} = 25^{\circ}\text{C} + 273.15 = 298 \text{ K}.$$

Example:**Converting between Temperature Scales: the Reaumur Scale**

The Reaumur scale is a temperature scale that was used widely in Europe in the 18th and 19th centuries. On the Reaumur temperature scale, the freezing point of water is 0°R and the boiling temperature is 80°R. If “room temperature” is 25°C on the Celsius scale, what is it on the Reaumur scale?

Strategy

To answer this question, we must compare the Reaumur scale to the Celsius scale. The difference between the freezing point and boiling point of water on the Reaumur scale is 80°R. On the Celsius scale it is 100°C. Therefore $100^{\circ}\text{C} = 80^{\circ}\text{R}$. Both scales start at 0° for freezing, so we can derive a simple formula to convert between temperatures on the two scales.

Solution

1. Derive a formula to convert from one scale to the other:

Equation:

$$T_{\text{°R}} = \frac{0.8^{\circ}\text{R}}{^{\circ}\text{C}} \times T_{\text{°C}}.$$

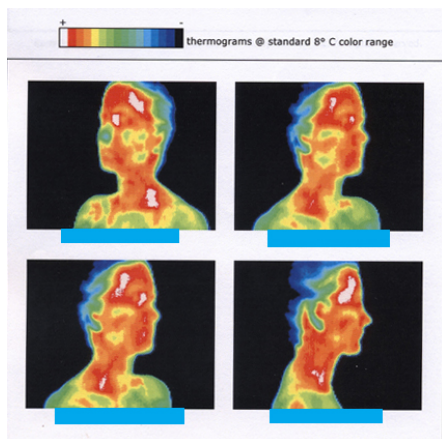
2. Plug the known value into the equation and solve:

Equation:

$$T_{\text{R}} = \frac{0.8^{\circ}\text{R}}{^{\circ}\text{C}} \times 25^{\circ}\text{C} = 20^{\circ}\text{R}.$$

Temperature Ranges in the Universe

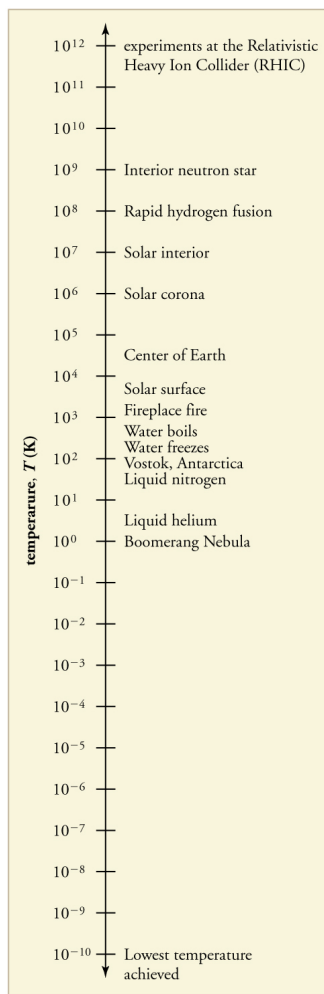
[\[link\]](#) shows the wide range of temperatures found in the universe. Human beings have been known to survive with body temperatures within a small range, from 24°C to 44°C (75°F to 111°F). The average normal body temperature is usually given as 37.0°C (98.6°F), and variations in this temperature can indicate a medical condition: a fever, an infection, a tumor, or circulatory problems (see [\[link\]](#)).



This image of radiation from a person's body (an infrared thermograph) shows the location of temperature abnormalities in the upper body. Dark blue corresponds to cold areas and red to white corresponds to hot areas. An elevated temperature might be an indication of malignant tissue (a cancerous tumor in the breast, for example), while a depressed temperature

might be due to a decline in blood flow from a clot. In this case, the abnormalities are caused by a condition called hyperhidrosis.
(credit: Porcelina81, Wikimedia Commons)

The lowest temperatures ever recorded have been measured during laboratory experiments: 4.5×10^{-10} K at the Massachusetts Institute of Technology (USA), and 1.0×10^{-10} K at Helsinki University of Technology (Finland). In comparison, the coldest recorded place on Earth's surface is Vostok, Antarctica at 183 K (-89°C), and the coldest place (outside the lab) known in the universe is the Boomerang Nebula, with a temperature of 1 K.

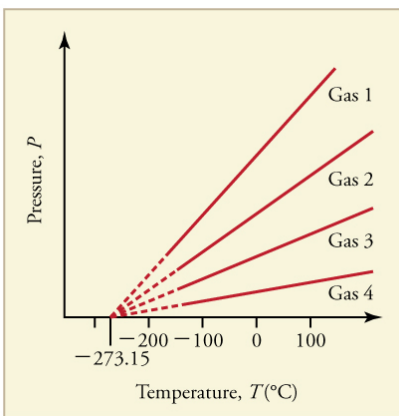


Each increment on this logarithmic scale indicates an increase by a factor of ten, and thus illustrates the tremendous range of temperatures in nature. Note that zero on a logarithmic scale would occur off the bottom of the page at infinity.

Note:

Making Connections: Absolute Zero

What is absolute zero? Absolute zero is the temperature at which all molecular motion has ceased. The concept of absolute zero arises from the behavior of gases. [\[link\]](#) shows how the pressure of gases at a constant volume decreases as temperature decreases. Various scientists have noted that the pressures of gases extrapolate to zero at the same temperature, -273.15°C . This extrapolation implies that there is a lowest temperature. This temperature is called *absolute zero*. Today we know that most gases first liquefy and then freeze, and it is not actually possible to reach absolute zero. The numerical value of absolute zero temperature is -273.15°C or 0 K.



Graph of pressure versus temperature for various

gases kept at a constant volume. Note that all of the graphs extrapolate to zero pressure at the same temperature.

Thermal Equilibrium and the Zeroth Law of Thermodynamics

Thermometers actually take their *own* temperature, not the temperature of the object they are measuring. This raises the question of how we can be certain that a thermometer measures the temperature of the object with which it is in contact. It is based on the fact that any two systems placed in *thermal contact* (meaning heat transfer can occur between them) will reach the same temperature. That is, heat will flow from the hotter object to the cooler one until they have exactly the same temperature. The objects are then in **thermal equilibrium**, and no further changes will occur. The systems interact and change because their temperatures differ, and the changes stop once their temperatures are the same. Thus, if enough time is allowed for this transfer of heat to run its course, the temperature a thermometer registers *does* represent the system with which it is in thermal equilibrium. Thermal equilibrium is established when two bodies are in contact with each other and can freely exchange energy.

Furthermore, experimentation has shown that if two systems, A and B, are in thermal equilibrium with each another, and B is in thermal equilibrium with a third system C, then A is also in thermal equilibrium with C. This conclusion may seem obvious, because all three have the same temperature, but it is basic to thermodynamics. It is called the **zeroth law of thermodynamics**.

Note:

The Zeroth Law of Thermodynamics

If two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

This law was postulated in the 1930s, after the first and second laws of thermodynamics had been developed and named. It is called the *zeroth law* because it comes logically before the first and second laws (discussed in [Thermodynamics](#)). An example of this law in action is seen in babies in incubators: babies in incubators normally have very few clothes on, so to an observer they look as if they may not be warm enough. However, the temperature of the air, the cot, and the baby is the same, because they are in thermal equilibrium, which is accomplished by maintaining air temperature to keep the baby comfortable.

Exercise:
Check Your Understanding

Problem: Does the temperature of a body depend on its size?

Solution:

No, the system can be divided into smaller parts each of which is at the same temperature. We say that the temperature is an *intensive* quantity. Intensive quantities are independent of size.

Section Summary

- Temperature is the quantity measured by a thermometer.
- Temperature is related to the average kinetic energy of atoms and molecules in a system.
- Absolute zero is the temperature at which there is no molecular motion.
- There are three main temperature scales: Celsius, Fahrenheit, and Kelvin.
- Temperatures on one scale can be converted to temperatures on another scale using the following equations:

Equation:

$$T_{\text{°F}} = \frac{9}{5}T_{\text{°C}} + 32$$

Equation:

$$T_{\text{°C}} = \frac{5}{9}(T_{\text{°F}} - 32)$$

Equation:

$$T_{\text{K}} = T_{\text{°C}} + 273.15$$

Equation:

$$T_{\text{°C}} = T_{\text{K}} - 273.15$$

- Systems are in thermal equilibrium when they have the same temperature.
- Thermal equilibrium occurs when two bodies are in contact with each other and can freely exchange energy.
- The zeroth law of thermodynamics states that when two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

Conceptual Questions

Exercise:

Problem: What does it mean to say that two systems are in thermal equilibrium?

Exercise:

Problem:

Give an example of a physical property that varies with temperature and describe how it is used to measure temperature.

Exercise:

Problem:

When a cold alcohol thermometer is placed in a hot liquid, the column of alcohol goes *down* slightly before going up. Explain why.

Exercise:

Problem:

If you add boiling water to a cup at room temperature, what would you expect the final equilibrium temperature of the unit to be? You will need to include the surroundings as part of the system. Consider the zeroth law of thermodynamics.

Problems & Exercises

Exercise:

Problem: What is the Fahrenheit temperature of a person with a 39.0°C fever?

Solution:

102°F

Exercise:

Problem:

Frost damage to most plants occurs at temperatures of 28.0°F or lower. What is this temperature on the Kelvin scale?

Exercise:

Problem:

To conserve energy, room temperatures are kept at 68.0°F in the winter and 78.0°F in the summer. What are these temperatures on the Celsius scale?

Solution:

20.0°C and 25.6°C

Exercise:**Problem:**

A tungsten light bulb filament may operate at 2900 K . What is its Fahrenheit temperature? What is this on the Celsius scale?

Exercise:**Problem:**

The surface temperature of the Sun is about 5750 K . What is this temperature on the Fahrenheit scale?

Solution:

9890°F

Exercise:**Problem:**

One of the hottest temperatures ever recorded on the surface of Earth was 134°F in Death Valley, CA. What is this temperature in Celsius degrees? What is this temperature in Kelvin?

Exercise:**Problem:**

(a) Suppose a cold front blows into your locale and drops the temperature by 40.0 Fahrenheit degrees. How many degrees Celsius does the temperature decrease when there is a 40.0°F decrease in temperature? (b) Show that any change in temperature in Fahrenheit degrees is nine-fifths the change in Celsius degrees.

Solution:

(a) 22.2°C

$$\begin{aligned}
 \Delta T(^{\circ}\text{F}) &= T_2(^{\circ}\text{F}) - T_1(^{\circ}\text{F}) \\
 \text{(b)} \quad &= \frac{9}{5}T_2(^{\circ}\text{C}) + 32.0^{\circ} - \left(\frac{9}{5}T_1(^{\circ}\text{C}) + 32.0^{\circ}\right) \\
 &= \frac{9}{5}(T_2(^{\circ}\text{C}) - T_1(^{\circ}\text{C})) = \frac{9}{5}\Delta T(^{\circ}\text{C})
 \end{aligned}$$

Exercise:

Problem:

(a) At what temperature do the Fahrenheit and Celsius scales have the same numerical value? (b) At what temperature do the Fahrenheit and Kelvin scales have the same numerical value?

Glossary

temperature

the quantity measured by a thermometer

Celsius scale

temperature scale in which the freezing point of water is 0°C and the boiling point of water is 100°C

degree Celsius

unit on the Celsius temperature scale

Fahrenheit scale

temperature scale in which the freezing point of water is 32°F and the boiling point of water is 212°F

degree Fahrenheit

unit on the Fahrenheit temperature scale

Kelvin scale

temperature scale in which 0 K is the lowest possible temperature, representing absolute zero

absolute zero

the lowest possible temperature; the temperature at which all molecular motion ceases

thermal equilibrium

the condition in which heat no longer flows between two objects that are in contact; the two objects have the same temperature

zeroth law of thermodynamics

law that states that if two objects are in thermal equilibrium, and a third object is in thermal equilibrium with one of those objects, it is also in thermal equilibrium with the other object

Thermal Expansion of Solids and Liquids

- Define and describe thermal expansion.
- Calculate the linear expansion of an object given its initial length, change in temperature, and coefficient of linear expansion.
- Calculate the volume expansion of an object given its initial volume, change in temperature, and coefficient of volume expansion.
- Calculate thermal stress on an object given its original volume, temperature change, volume change, and bulk modulus.



Thermal expansion joints like these in the Auckland Harbour Bridge in New Zealand allow bridges to change length without buckling. (credit: Ingolfson, Wikimedia Commons)

The expansion of alcohol in a thermometer is one of many commonly encountered examples of **thermal expansion**, the change in size or volume of a given mass with temperature. Hot air rises because its volume increases, which causes the hot air's density to be smaller than the density of surrounding air, causing a buoyant (upward) force on the hot air. The same happens in all liquids and gases, driving natural heat transfer upwards in homes, oceans, and weather systems. Solids also undergo thermal expansion. Railroad tracks and bridges, for example, have expansion joints to allow them to freely expand and contract with temperature changes.

What are the basic properties of thermal expansion? First, thermal expansion is clearly related to temperature change. The greater the temperature change, the more a bimetallic strip will bend. Second, it depends on the material. In a thermometer, for example, the expansion of alcohol is much greater than the expansion of the glass containing it.

What is the underlying cause of thermal expansion? As is discussed in [Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature](#), an increase in temperature implies an increase in the kinetic energy of the individual atoms. In a solid, unlike in a gas, the atoms or molecules are closely packed together, but their kinetic energy (in the form of small, rapid vibrations) pushes neighboring atoms or molecules apart from each other. This neighbor-to-neighbor pushing results in a slightly greater distance, on average, between neighbors, and adds up to a larger size for the whole body. For most substances under ordinary conditions, there is no preferred direction, and an increase in temperature will increase the solid's size by a certain fraction in each dimension.

Note:

Linear Thermal Expansion—Thermal Expansion in One Dimension

The change in length ΔL is proportional to length L . The dependence of thermal expansion on temperature, substance, and length is summarized in the equation

Equation:

$$\Delta L = \alpha L \Delta T,$$

where ΔL is the change in length L , ΔT is the change in temperature, and α is the **coefficient of linear expansion**, which varies slightly with temperature.

[\[link\]](#) lists representative values of the coefficient of linear expansion, which may have units of $1/^{\circ}\text{C}$ or $1/\text{K}$. Because the size of a kelvin and a degree Celsius are the same, both α and ΔT can be expressed in units of kelvins or degrees Celsius. The equation $\Delta L = \alpha L \Delta T$ is accurate for small changes in temperature and can be used for large changes in temperature if an average value of α is used.

Material	Coefficient of linear expansion $\alpha(1/^{\circ}\text{C})$	Coefficient of volume expansion $\beta(1/^{\circ}\text{C})$
Solids		
Aluminum	25×10^{-6}	75×10^{-6}
Brass	19×10^{-6}	56×10^{-6}
Copper	17×10^{-6}	51×10^{-6}

Material	Coefficient of linear expansion $\alpha(1/^{\circ}\text{C})$	Coefficient of volume expansion $\beta(1/^{\circ}\text{C})$
Gold	14×10^{-6}	42×10^{-6}
Iron or Steel	12×10^{-6}	35×10^{-6}
Invar (Nickel-iron alloy)	0.9×10^{-6}	2.7×10^{-6}
Lead	29×10^{-6}	87×10^{-6}
Silver	18×10^{-6}	54×10^{-6}
Glass (ordinary)	9×10^{-6}	27×10^{-6}
Glass (Pyrex®)	3×10^{-6}	9×10^{-6}
Quartz	0.4×10^{-6}	1×10^{-6}

Material	Coefficient of linear expansion	Coefficient of volume expansion
	$\alpha(1/^{\circ}\text{C})$	$\beta(1/^{\circ}\text{C})$
Concrete, Brick	$\sim 12 \times 10^{-6}$	$\sim 36 \times 10^{-6}$
Marble (average)	7×10^{-6}	2.1×10^{-5}
Liquids		
Ether		1650×10^{-6}
Ethyl alcohol		1100×10^{-6}
Petrol		950×10^{-6}
Glycerin		500×10^{-6}
Mercury		180×10^{-6}

Material	Coefficient of linear expansion $\alpha(1/^{\circ}\text{C})$	Coefficient of volume expansion $\beta(1/^{\circ}\text{C})$
Water		210×10^{-6}
Gases		
Air and most other gases at atmospheric pressure		3400×10^{-6}

Thermal Expansion Coefficients at 20°C[\[footnote\]](#)

Values for liquids and gases are approximate.

Example:

Calculating Linear Thermal Expansion: The Golden Gate Bridge

The main span of San Francisco's Golden Gate Bridge is 1275 m long at its coldest. The bridge is exposed to temperatures ranging from -15°C to 40°C . What is its change in length between these temperatures? Assume that the bridge is made entirely of steel.

Strategy

Use the equation for linear thermal expansion $\Delta L = \alpha L \Delta T$ to calculate the change in length, ΔL . Use the coefficient of linear expansion, α , for steel from [\[link\]](#), and note that the change in temperature, ΔT , is 55°C .

Solution

Plug all of the known values into the equation to solve for ΔL .

Equation:

$$\Delta L = \alpha L \Delta T = \left(\frac{12 \times 10^{-6}}{^{\circ}\text{C}} \right) (1275 \text{ m}) (55^{\circ}\text{C}) = 0.84 \text{ m}.$$

Discussion

Although not large compared with the length of the bridge, this change in length is observable. It is generally spread over many expansion joints so that the expansion at each joint is small.

Thermal Expansion in Two and Three Dimensions

Objects expand in all dimensions, as illustrated in [\[link\]](#). That is, their areas and volumes, as well as their lengths, increase with temperature. Holes also get larger with temperature. If you cut a hole in a metal plate, the remaining material will expand exactly as it would if the plug was still in place. The plug would get bigger, and so the hole must get bigger too. (Think of the ring of neighboring atoms or molecules on the wall of the hole as pushing each other farther apart as temperature increases. Obviously, the ring of neighbors must get slightly larger, so the hole gets slightly larger).

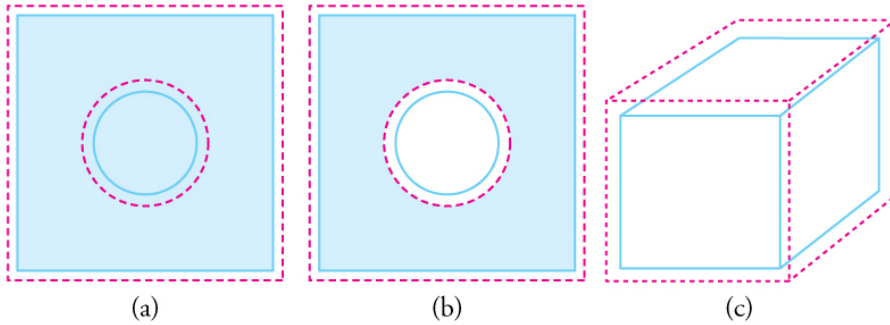
Note:**Thermal Expansion in Two Dimensions**

For small temperature changes, the change in area ΔA is given by

Equation:

$$\Delta A = 2\alpha A \Delta T,$$

where ΔA is the change in area A , ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.



In general, objects expand in all directions as temperature increases. In these drawings, the original boundaries of the objects are shown with solid lines, and the expanded boundaries with dashed lines. (a) Area increases because both length and width increase. The area of a circular plug also increases. (b) If the plug is removed, the hole it leaves becomes larger with increasing temperature, just as if the expanding plug were still in place. (c) Volume also increases, because all three dimensions increase.

Note:

Thermal Expansion in Three Dimensions

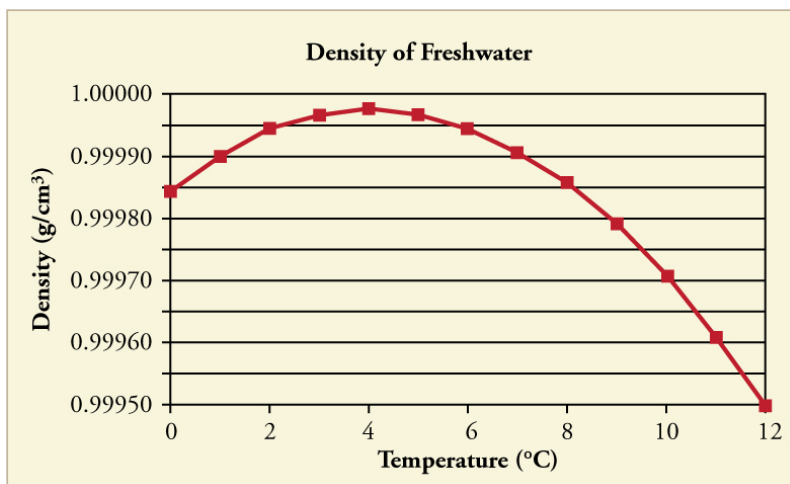
The change in volume ΔV is very nearly $\Delta V = 3\alpha V \Delta T$. This equation is usually written as

Equation:

$$\Delta V = \beta V \Delta T,$$

where β is the **coefficient of volume expansion** and $\beta \approx 3\alpha$. Note that the values of β in [\[link\]](#) are almost exactly equal to 3α .

In general, objects will expand with increasing temperature. Water is the most important exception to this rule. Water expands with increasing temperature (its density *decreases*) when it is at temperatures greater than 4°C (40°F). However, it expands with *decreasing* temperature when it is between $+4^{\circ}\text{C}$ and 0°C (40°F to 32°F). Water is densest at $+4^{\circ}\text{C}$. (See [\[link\]](#).) Perhaps the most striking effect of this phenomenon is the freezing of water in a pond. When water near the surface cools down to 4°C it is denser than the remaining water and thus will sink to the bottom. This “turnover” results in a layer of warmer water near the surface, which is then cooled. Eventually the pond has a uniform temperature of 4°C . If the temperature in the surface layer drops below 4°C , the water is less dense than the water below, and thus stays near the top. As a result, the pond surface can completely freeze over. The ice on top of liquid water provides an insulating layer from winter’s harsh exterior air temperatures. Fish and other aquatic life can survive in 4°C water beneath ice, due to this unusual characteristic of water. It also produces circulation of water in the pond that is necessary for a healthy ecosystem of the body of water.



The density of water as a function of temperature. Note that the thermal expansion is actually very small. The maximum density at $+4^{\circ}\text{C}$ is only 0.0075% greater than the density at 2°C , and 0.012% greater than that at 0°C .

Note:**Making Connections: Real-World Connections—Filling the Tank**

Differences in the thermal expansion of materials can lead to interesting effects at the gas station. One example is the dripping of gasoline from a freshly filled tank on a hot day. Gasoline starts out at the temperature of the ground under the gas station, which is cooler than the air temperature above. The gasoline cools the steel tank when it is filled. Both gasoline and steel tank expand as they warm to air temperature, but gasoline expands much more than steel, and so it may overflow.

This difference in expansion can also cause problems when interpreting the gasoline gauge. The actual amount (mass) of gasoline left in the tank when the gauge hits “empty” is a lot less in the summer than in the winter. The gasoline has the same volume as it does in the winter when the “add fuel” light goes on, but because the gasoline has expanded, there is less mass. If you are used to getting another 40 miles on “empty” in the winter, beware—you will probably run out much more quickly in the summer.



Because the gas expands more than the gas tank with increasing temperature, you can't drive as many miles on “empty” in the summer as you can in the winter.

(credit: Hector Alejandro,
Flickr)

Example:

Calculating Thermal Expansion: Gas vs. Gas Tank

Suppose your 60.0-L (15.9-gal) steel gasoline tank is full of gas, so both the tank and the gasoline have a temperature of 15.0°C. How much gasoline has spilled by the time they warm to 35.0°C?

Strategy

The tank and gasoline increase in volume, but the gasoline increases more, so the amount spilled is the difference in their volume changes. (The gasoline tank can be treated as solid steel.) We can use the equation for volume expansion to calculate the change in volume of the gasoline and of the tank.

Solution

1. Use the equation for volume expansion to calculate the increase in volume of the steel tank:

Equation:

$$\Delta V_s = \beta_s V_s \Delta T.$$

2. The increase in volume of the gasoline is given by this equation:

Equation:

$$\Delta V_{\text{gas}} = \beta_{\text{gas}} V_{\text{gas}} \Delta T.$$

3. Find the difference in volume to determine the amount spilled as

Equation:

$$V_{\text{spill}} = \Delta V_{\text{gas}} - \Delta V_s.$$

Alternatively, we can combine these three equations into a single equation. (Note that the original volumes are equal.)

Equation:

$$\begin{aligned}
 V_{\text{spill}} &= (\beta_{\text{gas}} - \beta_{\text{s}})V\Delta T \\
 &= [(950 - 35) \times 10^{-6} / ^\circ\text{C}] (60.0 \text{ L})(20.0^\circ\text{C}) \\
 &= 1.10 \text{ L.}
 \end{aligned}$$

Discussion

This amount is significant, particularly for a 60.0-L tank. The effect is so striking because the gasoline and steel expand quickly. The rate of change in thermal properties is discussed in [Heat and Heat Transfer Methods](#).

If you try to cap the tank tightly to prevent overflow, you will find that it leaks anyway, either around the cap or by bursting the tank. Tightly constricting the expanding gas is equivalent to compressing it, and both liquids and solids resist being compressed with extremely large forces. To avoid rupturing rigid containers, these containers have air gaps, which allow them to expand and contract without stressing them.

Thermal Stress

Thermal stress is created by thermal expansion or contraction (see [Elasticity: Stress and Strain](#) for a discussion of stress and strain). Thermal stress can be destructive, such as when expanding gasoline ruptures a tank. It can also be useful, for example, when two parts are joined together by heating one in manufacturing, then slipping it over the other and allowing the combination to cool. Thermal stress can explain many phenomena, such as the weathering of rocks and pavement by the expansion of ice when it freezes.

Example:

Calculating Thermal Stress: Gas Pressure

What pressure would be created in the gasoline tank considered in [\[link\]](#), if the gasoline increases in temperature from 15.0°C to 35.0°C without being allowed to expand? Assume that the bulk modulus B for gasoline is $1.00 \times 10^9 \text{ N/m}^2$. (For more on bulk modulus, see [Elasticity: Stress and Strain](#).)

Strategy

To solve this problem, we must use the following equation, which relates a change in volume ΔV to pressure:

Equation:

$$\Delta V = \frac{1}{B} \frac{F}{A} V_0,$$

where F/A is pressure, V_0 is the original volume, and B is the bulk modulus of the material involved. We will use the amount spilled in [\[link\]](#) as the change in volume, ΔV .

Solution

1. Rearrange the equation for calculating pressure:

Equation:

$$P = \frac{F}{A} = \frac{\Delta V}{V_0} B.$$

2. Insert the known values. The bulk modulus for gasoline is $B = 1.00 \times 10^9 \text{ N/m}^2$. In the previous example, the change in volume $\Delta V = 1.10 \text{ L}$ is the amount that would spill. Here, $V_0 = 60.0 \text{ L}$ is the original volume of the gasoline. Substituting these values into the equation, we obtain

Equation:

$$P = \frac{1.10 \text{ L}}{60.0 \text{ L}} (1.00 \times 10^9 \text{ Pa}) = 1.83 \times 10^7 \text{ Pa}.$$

Discussion

This pressure is about 2500 lb/in^2 , *much* more than a gasoline tank can handle.

Forces and pressures created by thermal stress are typically as great as that in the example above. Railroad tracks and roadways can buckle on hot days if they lack sufficient expansion joints. (See [\[link\]](#).) Power lines sag more in the summer than in the winter, and will snap in cold weather if there is

insufficient slack. Cracks open and close in plaster walls as a house warms and cools. Glass cooking pans will crack if cooled rapidly or unevenly, because of differential contraction and the stresses it creates. (Pyrex® is less susceptible because of its small coefficient of thermal expansion.) Nuclear reactor pressure vessels are threatened by overly rapid cooling, and although none have failed, several have been cooled faster than considered desirable. Biological cells are ruptured when foods are frozen, detracting from their taste. Repeated thawing and freezing accentuate the damage. Even the oceans can be affected. A significant portion of the rise in sea level that is resulting from global warming is due to the thermal expansion of sea water.



Thermal stress contributes to the formation of potholes.
(credit: Editor5807, Wikimedia Commons)

Metal is regularly used in the human body for hip and knee implants. Most implants need to be replaced over time because, among other things, metal does not bond with bone. Researchers are trying to find better metal coatings that would allow metal-to-bone bonding. One challenge is to find a coating that has an expansion coefficient similar to that of metal. If the

expansion coefficients are too different, the thermal stresses during the manufacturing process lead to cracks at the coating-metal interface.

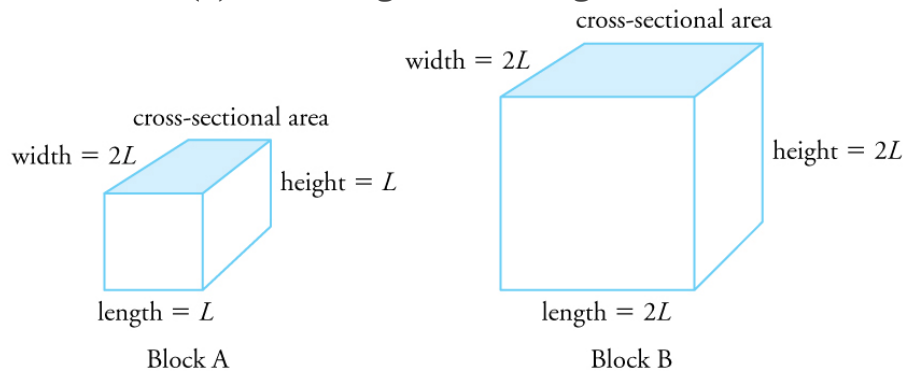
Another example of thermal stress is found in the mouth. Dental fillings can expand differently from tooth enamel. It can give pain when eating ice cream or having a hot drink. Cracks might occur in the filling. Metal fillings (gold, silver, etc.) are being replaced by composite fillings (porcelain), which have smaller coefficients of expansion, and are closer to those of teeth.

Exercise:

Check Your Understanding

Problem:

Two blocks, A and B, are made of the same material. Block A has dimensions $l \times w \times h = L \times 2L \times L$ and Block B has dimensions $2L \times 2L \times 2L$. If the temperature changes, what is (a) the change in the volume of the two blocks, (b) the change in the cross-sectional area $l \times w$, and (c) the change in the height h of the two blocks?



Solution:

(a) The change in volume is proportional to the original volume. Block A has a volume of $L \times 2L \times L = 2L^3$. Block B has a volume of $2L \times 2L \times 2L = 8L^3$, which is 4 times that of Block A. Thus the change in volume of Block B should be 4 times the change in volume of Block A.

(b) The change in area is proportional to the area. The cross-sectional area of Block A is $L \times 2L = 2L^2$, while that of Block B is

$2L \times 2L = 4L^2$. Because cross-sectional area of Block B is twice that of Block A, the change in the cross-sectional area of Block B is twice that of Block A.

(c) The change in height is proportional to the original height. Because the original height of Block B is twice that of A, the change in the height of Block B is twice that of Block A.

Section Summary

- Thermal expansion is the increase, or decrease, of the size (length, area, or volume) of a body due to a change in temperature.
- Thermal expansion is large for gases, and relatively small, but not negligible, for liquids and solids.
- Linear thermal expansion is

Equation:

$$\Delta L = \alpha L \Delta T,$$

where ΔL is the change in length L , ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.

- The change in area due to thermal expansion is

Equation:

$$\Delta A = 2\alpha A \Delta T,$$

where ΔA is the change in area.

- The change in volume due to thermal expansion is

Equation:

$$\Delta V = \beta V \Delta T,$$

where β is the coefficient of volume expansion and $\beta \approx 3\alpha$. Thermal stress is created when thermal expansion is constrained.

Conceptual Questions

Exercise:**Problem:**

Thermal stresses caused by uneven cooling can easily break glass cookware. Explain why Pyrex®, a glass with a small coefficient of linear expansion, is less susceptible.

Exercise:**Problem:**

Water expands significantly when it freezes: a volume increase of about 9% occurs. As a result of this expansion and because of the formation and growth of crystals as water freezes, anywhere from 10% to 30% of biological cells are burst when animal or plant material is frozen. Discuss the implications of this cell damage for the prospect of preserving human bodies by freezing so that they can be thawed at some future date when it is hoped that all diseases are curable.

Exercise:**Problem:**

One method of getting a tight fit, say of a metal peg in a hole in a metal block, is to manufacture the peg slightly larger than the hole. The peg is then inserted when at a different temperature than the block. Should the block be hotter or colder than the peg during insertion? Explain your answer.

Exercise:**Problem:**

Does it really help to run hot water over a tight metal lid on a glass jar before trying to open it? Explain your answer.

Exercise:

Problem:

Liquids and solids expand with increasing temperature, because the kinetic energy of a body's atoms and molecules increases. Explain why some materials *shrink* with increasing temperature.

Problems & Exercises**Exercise:****Problem:**

The height of the Washington Monument is measured to be 170 m on a day when the temperature is 35.0°C . What will its height be on a day when the temperature falls to -10.0°C ? Although the monument is made of limestone, assume that its thermal coefficient of expansion is the same as marble's.

Solution:

169.98 m

Exercise:**Problem:**

How much taller does the Eiffel Tower become at the end of a day when the temperature has increased by 15°C ? Its original height is 321 m and you can assume it is made of steel.

Exercise:**Problem:**

What is the change in length of a 3.00-cm-long column of mercury if its temperature changes from 37.0°C to 40.0°C , assuming the mercury is unconstrained?

Solution:

$$5.4 \times 10^{-6} \text{ m}$$

Exercise:**Problem:**

How large an expansion gap should be left between steel railroad rails if they may reach a maximum temperature 35.0°C greater than when they were laid? Their original length is 10.0 m.

Exercise:**Problem:**

You are looking to purchase a small piece of land in Hong Kong. The price is “only” \$60,000 per square meter! The land title says the dimensions are $20 \text{ m} \times 30 \text{ m}$. By how much would the total price change if you measured the parcel with a steel tape measure on a day when the temperature was 20°C above normal?

Solution:

Because the area gets smaller, the price of the land DECREASES by ~\$17,000.

Exercise:**Problem:**

Global warming will produce rising sea levels partly due to melting ice caps but also due to the expansion of water as average ocean temperatures rise. To get some idea of the size of this effect, calculate the change in length of a column of water 1.00 km high for a temperature increase of 1.00°C . Note that this calculation is only approximate because ocean warming is not uniform with depth.

Exercise:**Problem:**

Show that 60.0 L of gasoline originally at 15.0°C will expand to 61.1 L when it warms to 35.0°C , as claimed in [\[link\]](#).

Solution:

Equation:

$$\begin{aligned} V &= V_0 + \Delta V = V_0(1 + \beta\Delta T) \\ &= (60.00 \text{ L})[1 + (950 \times 10^{-6}/^{\circ}\text{C})(35.0^{\circ}\text{C} - 15.0^{\circ}\text{C})] \\ &= 61.1 \text{ L} \end{aligned}$$

Exercise:

Problem:

(a) Suppose a meter stick made of steel and one made of invar (an alloy of iron and nickel) are the same length at 0°C . What is their difference in length at 22.0°C ? (b) Repeat the calculation for two 30.0-m-long surveyor's tapes.

Exercise:

Problem:

(a) If a 500-mL glass beaker is filled to the brim with ethyl alcohol at a temperature of 5.00°C , how much will overflow when its temperature reaches 22.0°C ? (b) How much less water would overflow under the same conditions?

Solution:

(a) 9.35 mL

(b) 7.56 mL

Exercise:

Problem:

Most automobiles have a coolant reservoir to catch radiator fluid that may overflow when the engine is hot. A radiator is made of copper and is filled to its 16.0-L capacity when at 10.0°C. What volume of radiator fluid will overflow when the radiator and fluid reach their 95.0°C operating temperature, given that the fluid's volume coefficient of expansion is $\beta = 400 \times 10^{-6} / ^\circ\text{C}$? Note that this coefficient is approximate, because most car radiators have operating temperatures of greater than 95.0°C.

Exercise:**Problem:**

A physicist makes a cup of instant coffee and notices that, as the coffee cools, its level drops 3.00 mm in the glass cup. Show that this decrease cannot be due to thermal contraction by calculating the decrease in level if the 350 cm³ of coffee is in a 7.00-cm-diameter cup and decreases in temperature from 95.0°C to 45.0°C. (Most of the drop in level is actually due to escaping bubbles of air.)

Solution:

0.832 mm

Exercise:**Problem:**

(a) The density of water at 0°C is very nearly 1000 kg/m³ (it is actually 999.84 kg/m³), whereas the density of ice at 0°C is 917 kg/m³. Calculate the pressure necessary to keep ice from expanding when it freezes, neglecting the effect such a large pressure would have on the freezing temperature. (This problem gives you only an indication of how large the forces associated with freezing water might be.) (b) What are the implications of this result for biological cells that are frozen?

Exercise:**Problem:**

Show that $\beta \approx 3\alpha$, by calculating the change in volume ΔV of a cube with sides of length L .

Solution:

We know how the length changes with temperature: $\Delta L = \alpha L_0 \Delta T$. Also we know that the volume of a cube is related to its length by $V = L^3$, so the final volume is then $V = V_0 + \Delta V = (L_0 + \Delta L)^3$. Substituting for ΔL gives

Equation:

$$V = (L_0 + \alpha L_0 \Delta T)^3 = L_0^3 (1 + \alpha \Delta T)^3.$$

Now, because $\alpha \Delta T$ is small, we can use the binomial expansion:

Equation:

$$V \approx L_0^3 (1 + 3\alpha \Delta T) = L_0^3 + 3\alpha L_0^3 \Delta T.$$

So writing the length terms in terms of volumes gives $V = V_0 + \Delta V \approx V_0 + 3\alpha V_0 \Delta T$, and so

Equation:

$$\Delta V = \beta V_0 \Delta T \approx 3\alpha V_0 \Delta T, \text{ or } \beta \approx 3\alpha.$$

Glossary

thermal expansion

the change in size or volume of an object with change in temperature

coefficient of linear expansion

α , the change in length, per unit length, per 1°C change in temperature; a constant used in the calculation of linear expansion; the coefficient of linear expansion depends on the material and to some degree on the temperature of the material

coefficient of volume expansion

β , the change in volume, per unit volume, per 1°C change in temperature

thermal stress

stress caused by thermal expansion or contraction

The Ideal Gas Law

- State the ideal gas law in terms of molecules and in terms of moles.
- Use the ideal gas law to calculate pressure change, temperature change, volume change, or the number of molecules or moles in a given volume.
- Use Avogadro's number to convert between number of molecules and number of moles.

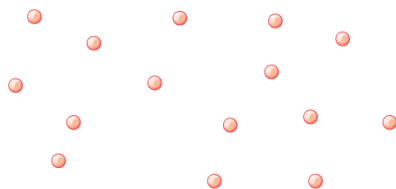


The air inside this hot air balloon flying over Putrajaya, Malaysia, is hotter than the ambient air. As a result, the balloon experiences a buoyant force pushing it upward.
(credit: Kevin Poh, Flickr)

In this section, we continue to explore the thermal behavior of gases. In particular, we examine the characteristics of atoms and molecules that compose gases. (Most gases, for example nitrogen, N_2 , and oxygen, O_2 , are composed of two or more atoms. We will primarily use the term “molecule” in discussing a gas because the term can also be applied to monatomic gases, such as helium.)

Gases are easily compressed. We can see evidence of this in [\[link\]](#), where you will note that gases have the *largest* coefficients of volume expansion. The large coefficients mean that gases expand and contract very rapidly with temperature changes. In addition, you will note that most gases expand at the *same* rate, or have the same β . This raises the question as to why gases should all act in nearly the same way, when liquids and solids have widely varying expansion rates.

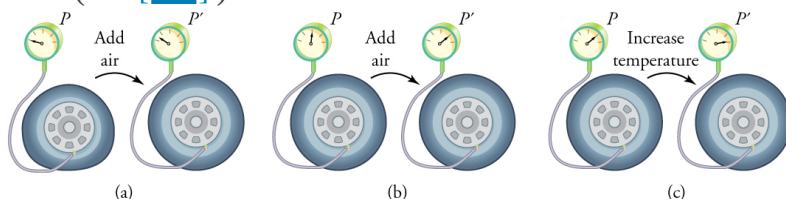
The answer lies in the large separation of atoms and molecules in gases, compared to their sizes, as illustrated in [\[link\]](#). Because atoms and molecules have large separations, forces between them can be ignored, except when they collide with each other during collisions. The motion of atoms and molecules (at temperatures well above the boiling temperature) is fast, such that the gas occupies all of the accessible volume and the expansion of gases is rapid. In contrast, in liquids and solids, atoms and molecules are closer together and are quite sensitive to the forces between them.



Atoms and molecules in a gas are typically widely separated, as shown.

Because the forces between them are quite weak at these distances, the properties of a gas depend more on the number of atoms per unit volume and on temperature than on the type of atom.

To get some idea of how pressure, temperature, and volume of a gas are related to one another, consider what happens when you pump air into an initially deflated tire. The tire's volume first increases in direct proportion to the amount of air injected, without much increase in the tire pressure. Once the tire has expanded to nearly its full size, the walls limit volume expansion. If we continue to pump air into it, the pressure increases. The pressure will further increase when the car is driven and the tires move. Most manufacturers specify optimal tire pressure for cold tires. (See [\[link\]](#).)



(a) When air is pumped into a deflated tire, its volume first increases without much increase in pressure. (b) When the tire is filled to a certain point, the tire walls resist further expansion and the pressure increases with

more air. (c) Once the tire is inflated, its pressure increases with temperature.

At room temperatures, collisions between atoms and molecules can be ignored. In this case, the gas is called an ideal gas, in which case the relationship between the pressure, volume, and temperature is given by the equation of state called the ideal gas law.

Note:

Ideal Gas Law

The **ideal gas law** states that

Equation:

$$PV = NkT,$$

where P is the absolute pressure of a gas, V is the volume it occupies, N is the number of atoms and molecules in the gas, and T is its absolute temperature. The constant k is called the **Boltzmann constant** in honor of Austrian physicist Ludwig Boltzmann (1844–1906) and has the value

Equation:

$$k = 1.38 \times 10^{-23} \text{ J/K}.$$

The ideal gas law can be derived from basic principles, but was originally deduced from experimental measurements of Charles' law (that volume occupied by a gas is proportional to temperature at a fixed pressure) and from Boyle's law (that for a fixed temperature, the product PV is a constant). In the ideal gas model, the volume occupied by its atoms and molecules is a negligible fraction of V . The ideal gas law describes the behavior of real gases under most conditions. (Note, for example, that N is the total number of atoms and molecules, independent of the type of gas.)

Let us see how the ideal gas law is consistent with the behavior of filling the tire when it is pumped slowly and the temperature is constant. At first, the pressure P is essentially equal to atmospheric pressure, and the volume V increases in direct proportion to the number of atoms and molecules N put into the tire. Once the volume of the tire is constant, the equation $PV = NkT$ predicts that the pressure should increase in proportion to *the number N of atoms and molecules*.

Example:

Calculating Pressure Changes Due to Temperature Changes: Tire Pressure

Suppose your bicycle tire is fully inflated, with an absolute pressure of 7.00×10^5 Pa (a gauge pressure of just under 90.0 lb/in²) at a temperature of 18.0°C. What is the pressure after its temperature has risen to 35.0°C? Assume that there are no appreciable leaks or changes in volume.

Strategy

The pressure in the tire is changing only because of changes in temperature. First we need to identify what we know and what we want to know, and then identify an equation to solve for the unknown.

We know the initial pressure $P_0 = 7.00 \times 10^5$ Pa, the initial temperature $T_0 = 18.0^\circ\text{C}$, and the final temperature $T_f = 35.0^\circ\text{C}$. We must find the final pressure P_f . How can we use the equation $PV = NkT$? At first, it may seem that not enough information is given, because the volume V and number of atoms N are not specified. What we can do is use the equation twice: $P_0V_0 = NkT_0$ and $P_fV_f = NkT_f$. If we divide P_fV_f by P_0V_0 we can come up with an equation that allows us to solve for P_f .

Equation:

$$\frac{P_f V_f}{P_0 V_0} = \frac{N_f k T_f}{N_0 k T_0}$$

Since the volume is constant, V_f and V_0 are the same and they cancel out. The same is true for N_f and N_0 , and k , which is a constant. Therefore,

Equation:

$$\frac{P_f}{P_0} = \frac{T_f}{T_0}.$$

We can then rearrange this to solve for P_f :

Equation:

$$P_f = P_0 \frac{T_f}{T_0},$$

where the temperature must be in units of kelvins, because T_0 and T_f are absolute temperatures.

Solution

1. Convert temperatures from Celsius to Kelvin.

Equation:

$$\begin{aligned} T_0 &= (18.0 + 273)\text{K} = 291 \text{ K} \\ T_f &= (35.0 + 273)\text{K} = 308 \text{ K} \end{aligned}$$

2. Substitute the known values into the equation.

Equation:

$$P_f = P_0 \frac{T_f}{T_0} = 7.00 \times 10^5 \text{ Pa} \left(\frac{308 \text{ K}}{291 \text{ K}} \right) = 7.41 \times 10^5 \text{ Pa}$$

Discussion

The final temperature is about 6% greater than the original temperature, so the final pressure is about 6% greater as well. Note that *absolute* pressure and *absolute* temperature must be used in the ideal gas law.

Note:

Making Connections: Take-Home Experiment—Refrigerating a Balloon

Inflate a balloon at room temperature. Leave the inflated balloon in the refrigerator overnight. What happens to the balloon, and why?

Example:

Calculating the Number of Molecules in a Cubic Meter of Gas

How many molecules are in a typical object, such as gas in a tire or water in a drink? We can use the ideal gas law to give us an idea of how large N typically is.

Calculate the number of molecules in a cubic meter of gas at standard temperature and pressure (STP), which is defined to be 0°C and atmospheric pressure.

Strategy

Because pressure, volume, and temperature are all specified, we can use the ideal gas law $PV = NkT$, to find N .

Solution

1. Identify the knowns.

Equation:

$$T = 0^\circ\text{C} = 273 \text{ K}$$

$$P = 1.01 \times 10^5 \text{ Pa}$$

$$V = 1.00 \text{ m}^3$$

$$k = 1.38 \times 10^{-23} \text{ J/K}$$

2. Identify the unknown: number of molecules, N .

3. Rearrange the ideal gas law to solve for N .

Equation:

$$PV = NkT$$

$$N = \frac{PV}{kT}$$

4. Substitute the known values into the equation and solve for N .

Equation:

$$N = \frac{PV}{kT} = \frac{(1.01 \times 10^5 \text{ Pa})(1.00 \text{ m}^3)}{(1.38 \times 10^{-23} \text{ J/K})(273 \text{ K})} = 2.68 \times 10^{25} \text{ molecules}$$

Discussion

This number is undeniably large, considering that a gas is mostly empty space. N is huge, even in small volumes. For example, 1 cm^3 of a gas at STP has 2.68×10^{19} molecules in it. Once again, note that N is the same for all types or mixtures of gases.

Moles and Avogadro's Number

It is sometimes convenient to work with a unit other than molecules when measuring the amount of substance. A **mole** (abbreviated mol) is defined to be the amount of a substance that contains as many atoms or molecules as there are atoms in exactly 12 grams (0.012 kg) of carbon-12. The actual number of atoms or molecules in one mole is called **Avogadro's number** (N_A), in recognition of Italian scientist Amedeo Avogadro (1776–1856). He developed the concept of the mole, based on the hypothesis that equal volumes of gas, at the same pressure and temperature, contain equal numbers of molecules. That is, the number is independent of the type of gas. This hypothesis has been confirmed, and the value of Avogadro's number is

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}.$$

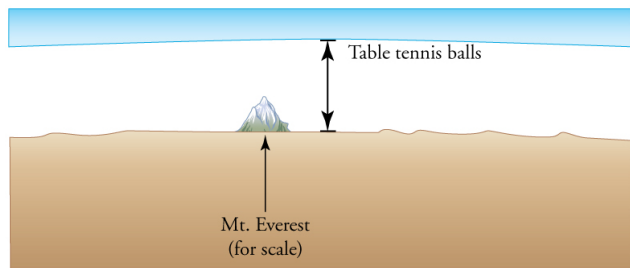
Note:

Avogadro's Number

One mole always contains 6.02×10^{23} particles (atoms or molecules), independent of the element or substance. A mole of any substance has a mass in grams equal to its molecular mass, which can be calculated from the atomic masses given in the periodic table of elements.

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}$$



How big is a mole? On a macroscopic level, one mole of table tennis balls would cover the Earth to a depth of about 40 km.

Exercise:
Check Your Understanding

Problem:

The active ingredient in a Tylenol pill is 325 mg of acetaminophen ($\text{C}_8\text{H}_9\text{NO}_2$). Find the number of active molecules of acetaminophen in a single pill.

Solution:

We first need to calculate the molar mass (the mass of one mole) of acetaminophen. To do this, we need to multiply the number of atoms of each element by the element's atomic mass.

Equation:

$$(8 \text{ moles of carbon})(12 \text{ grams/mole}) + (9 \text{ moles hydrogen})(1 \text{ gram/mole}) \\ + (1 \text{ mole nitrogen})(14 \text{ grams/mole}) + (2 \text{ moles oxygen})(16 \text{ grams/mole}) = 151 \text{ g}$$

Then we need to calculate the number of moles in 325 mg.

Equation:

$$\left(\frac{325 \text{ mg}}{151 \text{ grams/mole}} \right) \left(\frac{1 \text{ gram}}{1000 \text{ mg}} \right) = 2.15 \times 10^{-3} \text{ moles}$$

Then use Avogadro's number to calculate the number of molecules.

Equation:

$$N = (2.15 \times 10^{-3} \text{ moles}) (6.02 \times 10^{23} \text{ molecules/mole}) = 1.30 \times 10^{21} \text{ molecules}$$

Example:

Calculating Moles per Cubic Meter and Liters per Mole

Calculate: (a) the number of moles in 1.00 m^3 of gas at STP, and (b) the number of liters of gas per mole.

Strategy and Solution

(a) We are asked to find the number of moles per cubic meter, and we know from [\[link\]](#) that the number of molecules per cubic meter at STP is 2.68×10^{25} . The number of moles can be found by dividing the number of molecules by Avogadro's number. We let n stand for the number of moles,

Equation:

$$n \text{ mol/m}^3 = \frac{N \text{ molecules/m}^3}{6.02 \times 10^{23} \text{ molecules/mol}} = \frac{2.68 \times 10^{25} \text{ molecules/m}^3}{6.02 \times 10^{23} \text{ molecules/mol}} = 44.5 \text{ mol/m}^3.$$

(b) Using the value obtained for the number of moles in a cubic meter, and converting cubic meters to liters, we obtain

Equation:

$$\frac{(10^3 \text{ L/m}^3)}{44.5 \text{ mol/m}^3} = 22.5 \text{ L/mol.}$$

Discussion

This value is very close to the accepted value of 22.4 L/mol. The slight difference is due to rounding errors caused by using three-digit input. Again this number is the same for all gases. In other words, it is independent of the gas.

The (average) molar weight of air (approximately 80% N₂ and 20% O₂) is $M = 28.8 \text{ g}$. Thus the mass of one cubic meter of air is 1.28 kg. If a living room has dimensions $5 \text{ m} \times 5 \text{ m} \times 3 \text{ m}$, the mass of air inside the room is 96 kg, which is the typical mass of a human.

Exercise:

Check Your Understanding

Problem:

The density of air at standard conditions ($P = 1 \text{ atm}$ and $T = 20^\circ\text{C}$) is 1.28 kg/m^3 . At what pressure is the density 0.64 kg/m^3 if the temperature and number of molecules are kept constant?

Solution:

The best way to approach this question is to think about what is happening. If the density drops to half its original value and no molecules are lost, then the volume must double. If we look at the equation $PV = NkT$, we see that when the temperature is constant, the pressure is inversely proportional to volume. Therefore, if the volume doubles, the pressure must drop to half its original value, and $P_f = 0.50 \text{ atm}$.

The Ideal Gas Law Restated Using Moles

A very common expression of the ideal gas law uses the number of moles, n , rather than the number of atoms and molecules, N . We start from the ideal gas law,

Equation:

$$PV = NkT,$$

and multiply and divide the equation by Avogadro's number N_A . This gives

Equation:

$$PV = \frac{N}{N_A} N_A k T.$$

Note that $n = N/N_A$ is the number of moles. We define the universal gas constant $R = N_A k$, and obtain the ideal gas law in terms of moles.

Note:

Ideal Gas Law (in terms of moles)

The ideal gas law (in terms of moles) is

Equation:

$$PV = nRT.$$

The numerical value of R in SI units is

Equation:

$$R = N_A k = (6.02 \times 10^{23} \text{ mol}^{-1})(1.38 \times 10^{-23} \text{ J/K}) = 8.31 \text{ J/mol} \cdot \text{K}.$$

In other units,

Equation:

$$R = 1.99 \text{ cal/mol} \cdot \text{K}$$

$$R = 0.0821 \text{ L} \cdot \text{atm/mol} \cdot \text{K}.$$

You can use whichever value of R is most convenient for a particular problem.

Example:

Calculating Number of Moles: Gas in a Bike Tire

How many moles of gas are in a bike tire with a volume of $2.00 \times 10^{-3} \text{ m}^3$ (2.00 L), a pressure of $7.00 \times 10^5 \text{ Pa}$ (a gauge pressure of just under 90.0 lb/in²), and at a temperature of 18.0°C?

Strategy

Identify the knowns and unknowns, and choose an equation to solve for the unknown. In this case, we solve the ideal gas law, $PV = nRT$, for the number of moles n .

Solution

1. Identify the knowns.

Equation:

$$P = 7.00 \times 10^5 \text{ Pa}$$

$$V = 2.00 \times 10^{-3} \text{ m}^3$$

$$T = 18.0^\circ\text{C} = 291 \text{ K}$$

$$R = 8.31 \text{ J/mol} \cdot \text{K}$$

2. Rearrange the equation to solve for n and substitute known values.

Equation:

$$\begin{aligned} n &= \frac{PV}{RT} = \frac{(7.00 \times 10^5 \text{ Pa})(2.00 \times 10^{-3} \text{ m}^3)}{(8.31 \text{ J/mol}\cdot\text{K})(291 \text{ K})} \\ &= 0.579 \text{ mol} \end{aligned}$$

Discussion

The most convenient choice for R in this case is $8.31 \text{ J/mol}\cdot\text{K}$, because our known quantities are in SI units. The pressure and temperature are obtained from the initial conditions in [\[link\]](#), but we would get the same answer if we used the final values.

The ideal gas law can be considered to be another manifestation of the law of conservation of energy (see [Conservation of Energy](#)). Work done on a gas results in an increase in its energy, increasing pressure and/or temperature, or decreasing volume. This increased energy can also be viewed as increased internal kinetic energy, given the gas's atoms and molecules.

The Ideal Gas Law and Energy

Let us now examine the role of energy in the behavior of gases. When you inflate a bike tire by hand, you do work by repeatedly exerting a force through a distance. This energy goes into increasing the pressure of air inside the tire and increasing the temperature of the pump and the air.

The ideal gas law is closely related to energy: the units on both sides are joules. The right-hand side of the ideal gas law in $PV = NkT$ is NkT . This term is roughly the amount of translational kinetic energy of N atoms or molecules at an absolute temperature T , as we shall see formally in [Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature](#). The left-hand side of the ideal gas law is PV , which also has the units of joules. We know from our study of fluids that pressure is one type of potential energy per unit volume, so pressure multiplied by volume is energy. The important point is that there is energy in a gas related to both its pressure and its volume. The energy can be changed when the gas is doing work as it expands—something we explore in [Heat and Heat Transfer Methods](#)—similar to what occurs in gasoline or steam engines and turbines.

Note:

Problem-Solving Strategy: The Ideal Gas Law

Step 1 Examine the situation to determine that an ideal gas is involved. Most gases are nearly ideal.

Step 2 Make a list of what quantities are given, or can be inferred from the problem as stated (identify the known quantities). Convert known values into proper SI units (K for temperature, Pa for pressure, m^3 for volume, molecules for N , and moles for n).

Step 3 Identify exactly what needs to be determined in the problem (identify the unknown quantities). A written list is useful.

Step 4 Determine whether the number of molecules or the number of moles is known, in order to decide which form of the ideal gas law to use. The first form is $PV = NkT$ and involves N , the number of atoms or molecules. The second form is $PV = nRT$ and involves n , the number of moles.

Step 5 Solve the ideal gas law for the quantity to be determined (the unknown quantity). You may need to take a ratio of final states to initial states to eliminate the unknown quantities that are kept fixed.

Step 6 Substitute the known quantities, along with their units, into the appropriate equation, and obtain numerical solutions complete with units. Be certain to use absolute temperature and absolute pressure.

Step 7 Check the answer to see if it is reasonable: Does it make sense?

Exercise:

Check Your Understanding

Problem:

Liquids and solids have densities about 1000 times greater than gases. Explain how this implies that the distances between atoms and molecules in gases are about 10 times greater than the size of their atoms and molecules.

Solution:

Atoms and molecules are close together in solids and liquids. In gases they are separated by empty space. Thus gases have lower densities than liquids and solids. Density is mass per unit volume, and volume is related to the size of a body (such as a sphere) cubed. So if the distance between atoms and molecules increases by a factor of 10, then the volume occupied increases by a factor of 1000, and the density decreases by a factor of 1000.

Section Summary

- The ideal gas law relates the pressure and volume of a gas to the number of gas molecules and the temperature of the gas.
- The ideal gas law can be written in terms of the number of molecules of gas:

Equation:

$$PV = NkT,$$

where P is pressure, V is volume, T is temperature, N is number of molecules, and k is the Boltzmann constant

Equation:

$$k = 1.38 \times 10^{-23} \text{ J/K}.$$

- A mole is the number of atoms in a 12-g sample of carbon-12.
- The number of molecules in a mole is called Avogadro's number N_A ,

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}.$$

- A mole of any substance has a mass in grams equal to its molecular weight, which can be determined from the periodic table of elements.
- The ideal gas law can also be written and solved in terms of the number of moles of gas:

Equation:

$$PV = nRT,$$

where n is number of moles and R is the universal gas constant,

Equation:

$$R = 8.31 \text{ J/mol} \cdot \text{K}.$$

- The ideal gas law is generally valid at temperatures well above the boiling temperature.

Conceptual Questions

Exercise:

Problem:

Find out the human population of Earth. Is there a mole of people inhabiting Earth? If the average mass of a person is 60 kg, calculate the mass of a mole of people. How does the mass of a mole of people compare with the mass of Earth?

Exercise:

Problem:

Under what circumstances would you expect a gas to behave significantly differently than predicted by the ideal gas law?

Exercise:

Problem:

A constant-volume gas thermometer contains a fixed amount of gas. What property of the gas is measured to indicate its temperature?

Problems & Exercises

Exercise:

Problem:

The gauge pressure in your car tires is $2.50 \times 10^5 \text{ N/m}^2$ at a temperature of 35.0°C when you drive it onto a ferry boat to Alaska. What is their gauge pressure later, when their temperature has dropped to -40.0°C ?

Solution:

1.62 atm

Exercise:**Problem:**

Convert an absolute pressure of $7.00 \times 10^5 \text{ N/m}^2$ to gauge pressure in lb/in^2 . (This value was stated to be just less than 90.0 lb/in^2 in [\[link\]](#). Is it?)

Exercise:**Problem:**

Suppose a gas-filled incandescent light bulb is manufactured so that the gas inside the bulb is at atmospheric pressure when the bulb has a temperature of 20.0°C . (a) Find the gauge pressure inside such a bulb when it is hot, assuming its average temperature is 60.0°C (an approximation) and neglecting any change in volume due to thermal expansion or gas leaks. (b) The actual final pressure for the light bulb will be less than calculated in part (a) because the glass bulb will expand. What will the actual final pressure be, taking this into account? Is this a negligible difference?

Solution:

(a) 0.136 atm

(b) 0.135 atm. The difference between this value and the value from part (a) is negligible.

Exercise:**Problem:**

Large helium-filled balloons are used to lift scientific equipment to high altitudes. (a) What is the pressure inside such a balloon if it starts out at sea level with a temperature of 10.0°C and rises to an altitude where its volume is twenty times the original volume and its temperature is -50.0°C ? (b) What is the gauge pressure? (Assume atmospheric pressure is constant.)

Exercise:**Problem:**

Confirm that the units of nRT are those of energy for each value of R : (a) $8.31 \text{ J/mol} \cdot \text{K}$, (b) $1.99 \text{ cal/mol} \cdot \text{K}$, and (c) $0.0821 \text{ L} \cdot \text{atm/mol} \cdot \text{K}$.

Solution:

(a) $nRT = (\text{mol})(\text{J/mol} \cdot \text{K})(\text{K}) = \text{J}$

(b) $nRT = (\text{mol})(\text{cal/mol} \cdot \text{K})(\text{K}) = \text{cal}$

$$\begin{aligned}
 nRT &= (\text{mol})(\text{L} \cdot \text{atm}/\text{mol} \cdot \text{K})(\text{K}) \\
 \text{(c)} \quad &= \text{L} \cdot \text{atm} = (\text{m}^3)(\text{N}/\text{m}^2) \\
 &= \text{N} \cdot \text{m} = \text{J}
 \end{aligned}$$

Exercise:

Problem:

In the text, it was shown that $N/V = 2.68 \times 10^{25} \text{ m}^{-3}$ for gas at STP. (a) Show that this quantity is equivalent to $N/V = 2.68 \times 10^{19} \text{ cm}^{-3}$, as stated. (b) About how many atoms are there in one μm^3 (a cubic micrometer) at STP? (c) What does your answer to part (b) imply about the separation of atoms and molecules?

Exercise:

Problem:

Calculate the number of moles in the 2.00-L volume of air in the lungs of the average person. Note that the air is at 37.0°C (body temperature).

Solution:

$$7.86 \times 10^{-2} \text{ mol}$$

Exercise:

Problem:

An airplane passenger has 100 cm^3 of air in his stomach just before the plane takes off from a sea-level airport. What volume will the air have at cruising altitude if cabin pressure drops to $7.50 \times 10^4 \text{ N}/\text{m}^2$?

Exercise:

Problem:

(a) What is the volume (in km^3) of Avogadro's number of sand grains if each grain is a cube and has sides that are 1.0 mm long? (b) How many kilometers of beaches in length would this cover if the beach averages 100 m in width and 10.0 m in depth? Neglect air spaces between grains.

Solution:

$$\text{(a)} \quad 6.02 \times 10^5 \text{ km}^3$$

$$\text{(b)} \quad 6.02 \times 10^8 \text{ km}$$

Exercise:

Problem:

An expensive vacuum system can achieve a pressure as low as $1.00 \times 10^{-7} \text{ N/m}^2$ at 20°C . How many atoms are there in a cubic centimeter at this pressure and temperature?

Exercise:**Problem:**

The number density of gas atoms at a certain location in the space above our planet is about $1.00 \times 10^{11} \text{ m}^{-3}$, and the pressure is $2.75 \times 10^{-10} \text{ N/m}^2$ in this space. What is the temperature there?

Solution:

-73.9°C

Exercise:**Problem:**

A bicycle tire has a pressure of $7.00 \times 10^5 \text{ N/m}^2$ at a temperature of 18.0°C and contains 2.00 L of gas. What will its pressure be if you let out an amount of air that has a volume of 100 cm^3 at atmospheric pressure? Assume tire temperature and volume remain constant.

Exercise:**Problem:**

A high-pressure gas cylinder contains 50.0 L of toxic gas at a pressure of $1.40 \times 10^7 \text{ N/m}^2$ and a temperature of 25.0°C . Its valve leaks after the cylinder is dropped. The cylinder is cooled to dry ice temperature (-78.5°C) to reduce the leak rate and pressure so that it can be safely repaired. (a) What is the final pressure in the tank, assuming a negligible amount of gas leaks while being cooled and that there is no phase change? (b) What is the final pressure if one-tenth of the gas escapes? (c) To what temperature must the tank be cooled to reduce the pressure to 1.00 atm (assuming the gas does not change phase and that there is no leakage during cooling)? (d) Does cooling the tank appear to be a practical solution?

Solution:

(a) $9.14 \times 10^6 \text{ N/m}^2$

(b) $8.23 \times 10^6 \text{ N/m}^2$

(c) 2.16 K

(d) No. The final temperature needed is much too low to be easily achieved for a large object.

Exercise:

Problem:

Find the number of moles in 2.00 L of gas at 35.0°C and under $7.41 \times 10^7 \text{ N/m}^2$ of pressure.

Exercise:**Problem:**

Calculate the depth to which Avogadro's number of table tennis balls would cover Earth. Each ball has a diameter of 3.75 cm. Assume the space between balls adds an extra 25.0% to their volume and assume they are not crushed by their own weight.

Solution:

41 km

Exercise:**Problem:**

(a) What is the gauge pressure in a 25.0°C car tire containing 3.60 mol of gas in a 30.0 L volume? (b) What will its gauge pressure be if you add 1.00 L of gas originally at atmospheric pressure and 25.0°C? Assume the temperature returns to 25.0°C and the volume remains constant.

Exercise:**Problem:**

(a) In the deep space between galaxies, the density of atoms is as low as 10^6 atoms/m^3 , and the temperature is a frigid 2.7 K. What is the pressure? (b) What volume (in m^3) is occupied by 1 mol of gas? (c) If this volume is a cube, what is the length of its sides in kilometers?

Solution:

(a) $3.7 \times 10^{-17} \text{ Pa}$

(b) $6.0 \times 10^{17} \text{ m}^3$

(c) $8.4 \times 10^2 \text{ km}$

Glossary

ideal gas law

the physical law that relates the pressure and volume of a gas to the number of gas molecules or number of moles of gas and the temperature of the gas

Boltzmann constant

k , a physical constant that relates energy to temperature; $k = 1.38 \times 10^{-23} \text{ J/K}$

Avogadro's number

N_A , the number of molecules or atoms in one mole of a substance; $N_A = 6.02 \times 10^{23}$
particles/mole

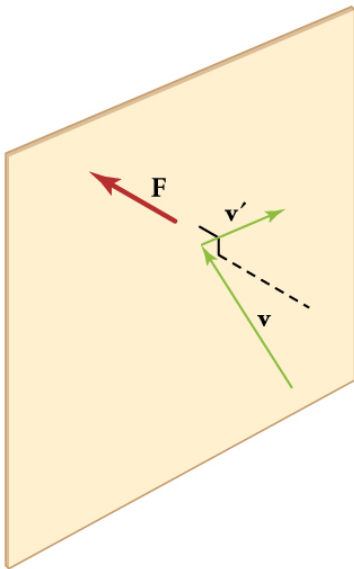
mole

the quantity of a substance whose mass (in grams) is equal to its molecular mass

Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature

- Express the ideal gas law in terms of molecular mass and velocity.
- Define thermal energy.
- Calculate the kinetic energy of a gas molecule, given its temperature.
- Describe the relationship between the temperature of a gas and the kinetic energy of atoms and molecules.
- Describe the distribution of speeds of molecules in a gas.

We have developed macroscopic definitions of pressure and temperature. Pressure is the force divided by the area on which the force is exerted, and temperature is measured with a thermometer. We gain a better understanding of pressure and temperature from the kinetic theory of gases, which assumes that atoms and molecules are in continuous random motion.



When a molecule collides with a rigid wall, the component of its momentum perpendicular to the wall is reversed. A force is thus exerted on the wall, creating pressure.

[\[link\]](#) shows an elastic collision of a gas molecule with the wall of a container, so that it exerts a force on the wall (by Newton's third law). Because a huge number of molecules will collide with the wall in a short time, we observe an average force per unit area. These collisions are the source of pressure in a gas. As the number of molecules increases, the number of collisions and thus the pressure increase. Similarly, the gas pressure is higher if the average velocity of molecules is higher. The actual relationship is derived in the [Things Great and Small](#) feature below. The following relationship is found:

Equation:

$$PV = \frac{1}{3}Nm\overline{v^2},$$

where P is the pressure (average force per unit area), V is the volume of gas in the container, N is the number of molecules in the container, m is the mass of a molecule, and $\overline{v^2}$ is the average of the molecular speed squared.

What can we learn from this atomic and molecular version of the ideal gas law? We can derive a relationship between temperature and the average translational kinetic energy of molecules in a gas. Recall the previous expression of the ideal gas law:

Equation:

$$PV = NkT.$$

Equating the right-hand side of this equation with the right-hand side of $PV = \frac{1}{3}Nm\overline{v^2}$ gives

Equation:

$$\frac{1}{3}Nm\overline{v^2} = NkT.$$

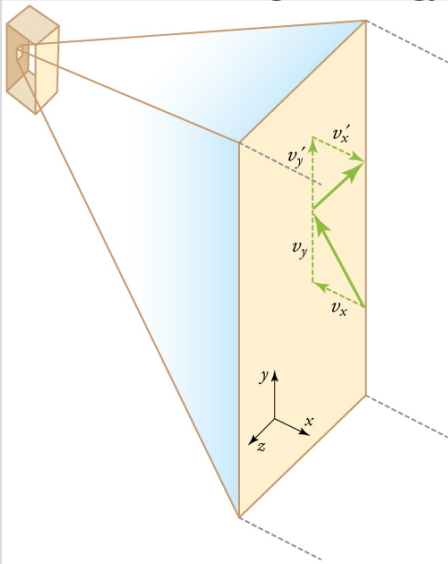
Note:

Making Connections: Things Great and Small—Atomic and Molecular Origin of Pressure in a Gas

[\[link\]](#) shows a box filled with a gas. We know from our previous discussions that putting more gas into the box produces greater pressure, and that increasing the temperature of the gas also produces a greater pressure. But why should increasing the temperature of the gas increase the pressure in the box? A look at the atomic and

molecular scale gives us some answers, and an alternative expression for the ideal gas law.

The figure shows an expanded view of an elastic collision of a gas molecule with the wall of a container. Calculating the average force exerted by such molecules will lead us to the ideal gas law, and to the connection between temperature and molecular kinetic energy. We assume that a molecule is small compared with the separation of molecules in the gas, and that its interaction with other molecules can be ignored. We also assume the wall is rigid and that the molecule's direction changes, but that its speed remains constant (and hence its kinetic energy and the magnitude of its momentum remain constant as well). This assumption is not always valid, but the same result is obtained with a more detailed description of the molecule's exchange of energy and momentum with the wall.



Gas in a box exerts an outward pressure on its walls. A molecule colliding with a rigid wall has the direction of its velocity and momentum in the x -direction reversed. This direction is perpendicular to the wall.

The components of its velocity momentum in the y - and z -directions are not changed, which means there is no force parallel to the wall.

If the molecule's velocity changes in the x -direction, its momentum changes from $-mv_x$ to $+mv_x$. Thus, its change in momentum is $\Delta mv = +mv_x - (-mv_x) = 2mv_x$. The force exerted on the molecule is given by

Equation:

$$F = \frac{\Delta p}{\Delta t} = \frac{2mv_x}{\Delta t}.$$

There is no force between the wall and the molecule until the molecule hits the wall. During the short time of the collision, the force between the molecule and wall is relatively large. We are looking for an average force; we take Δt to be the average time between collisions of the molecule with this wall. It is the time it would take the molecule to go across the box and back (a distance $2l$) at a speed of v_x . Thus $\Delta t = 2l/v_x$, and the expression for the force becomes

Equation:

$$F = \frac{2mv_x}{2l/v_x} = \frac{mv_x^2}{l}.$$

This force is due to *one* molecule. We multiply by the number of molecules N and use their average squared velocity to find the force

Equation:

$$F = N \frac{\overline{mv_x^2}}{l},$$

where the bar over a quantity means its average value. We would like to have the force in terms of the speed v , rather than the x -component of the velocity. We note that the total velocity squared is the sum of the squares of its components, so that

Equation:

$$\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}.$$

Because the velocities are random, their average components in all directions are the same:

Equation:

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2}.$$

Thus,

Equation:

$$\overline{v^2} = 3\overline{v_x^2},$$

or

Equation:

$$\overline{v_x^2} = \frac{1}{3}\overline{v^2}.$$

Substituting $\frac{1}{3}\overline{v^2}$ into the expression for F gives

Equation:

$$F = N \frac{m\overline{v^2}}{3l}.$$

The pressure is F/A , so that we obtain

Equation:

$$P = \frac{F}{A} = N \frac{m\overline{v^2}}{3Al} = \frac{1}{3} \frac{Nm\overline{v^2}}{V},$$

where we used $V = Al$ for the volume. This gives the important result.

Equation:

$$PV = \frac{1}{3}Nm\overline{v^2}$$

This equation is another expression of the ideal gas law.

We can get the average kinetic energy of a molecule, $\frac{1}{2}mv^2$, from the right-hand side of the equation by canceling N and multiplying by $3/2$. This calculation produces the result that the average kinetic energy of a molecule is directly related to absolute temperature.

Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

The average translational kinetic energy of a molecule, \overline{KE} , is called **thermal energy**. The equation $\overline{KE} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$ is a molecular interpretation of temperature, and it has been found to be valid for gases and reasonably accurate in liquids and solids. It is another definition of temperature based on an expression of the molecular energy.

It is sometimes useful to rearrange $\overline{KE} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$, and solve for the average speed of molecules in a gas in terms of temperature,

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}},$$

where v_{rms} stands for root-mean-square (rms) speed.

Example:

Calculating Kinetic Energy and Speed of a Gas Molecule

(a) What is the average kinetic energy of a gas molecule at 20.0°C (room temperature)? (b) Find the rms speed of a nitrogen molecule (N_2) at this temperature.

Strategy for (a)

The known in the equation for the average kinetic energy is the temperature.

Equation:

$$\overline{KE} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

Before substituting values into this equation, we must convert the given temperature to kelvins. This conversion gives $T = (20.0 + 273) \text{ K} = 293 \text{ K}$.

Solution for (a)

The temperature alone is sufficient to find the average translational kinetic energy. Substituting the temperature into the translational kinetic energy equation gives

Equation:

$$\overline{KE} = \frac{3}{2}kT = \frac{3}{2}(1.38 \times 10^{-23} \text{ J/K})(293 \text{ K}) = 6.07 \times 10^{-21} \text{ J}.$$

Strategy for (b)

Finding the rms speed of a nitrogen molecule involves a straightforward calculation using the equation

Equation:

$$\sqrt{v^2} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}},$$

but we must first find the mass of a nitrogen molecule. Using the molecular mass of nitrogen N_2 from the periodic table,

Equation:

$$m = \frac{2(14.0067) \times 10^{-3} \text{ kg/mol}}{6.02 \times 10^{23} \text{ mol}^{-1}} = 4.65 \times 10^{-26} \text{ kg}.$$

Solution for (b)

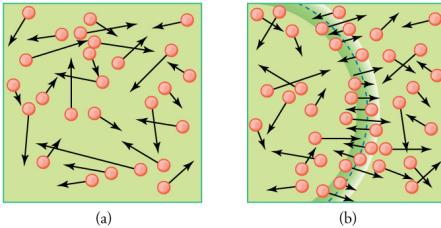
Substituting this mass and the value for k into the equation for v_{rms} yields

Equation:

$$v_{\text{rms}} = \sqrt{\frac{3kT}{m}} = \sqrt{\frac{3(1.38 \times 10^{-23} \text{ J/K})(293 \text{ K})}{4.65 \times 10^{-26} \text{ kg}}} = 511 \text{ m/s}.$$

Discussion

Note that the average kinetic energy of the molecule is independent of the type of molecule. The average translational kinetic energy depends only on absolute temperature. The kinetic energy is very small compared to macroscopic energies, so that we do not feel when an air molecule is hitting our skin. The rms velocity of the nitrogen molecule is surprisingly large. These large molecular velocities do not yield macroscopic movement of air, since the molecules move in all directions with equal likelihood. The *mean free path* (the distance a molecule can move on average between collisions) of molecules in air is very small, and so the molecules move rapidly but do not get very far in a second. The high value for rms speed is reflected in the speed of sound, however, which is about 340 m/s at room temperature. The faster the rms speed of air molecules, the faster that sound vibrations can be transferred through the air. The speed of sound increases with temperature and is greater in gases with small molecular masses, such as helium. (See [\[link\]](#).)



(a) There are many molecules moving so fast in an ordinary gas that they collide a billion times every second. (b) Individual molecules do not move very far in a small amount of time, but disturbances like sound waves are transmitted at speeds related to the molecular speeds.

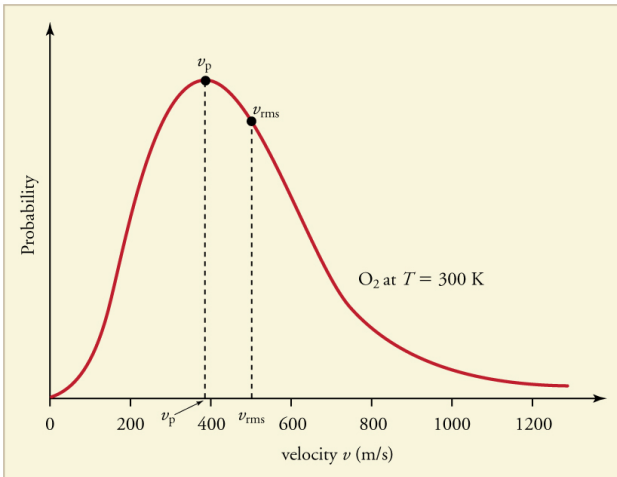
Note:

Making Connections: Historical Note—Kinetic Theory of Gases

The kinetic theory of gases was developed by Daniel Bernoulli (1700–1782), who is best known in physics for his work on fluid flow (hydrodynamics). Bernoulli’s work predates the atomistic view of matter established by Dalton.

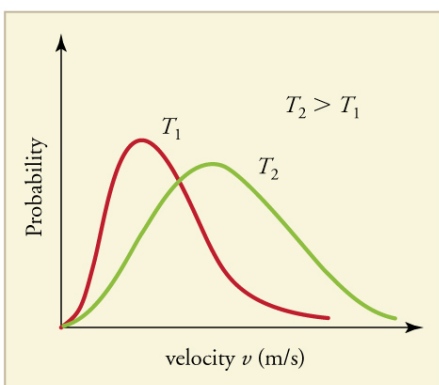
Distribution of Molecular Speeds

The motion of molecules in a gas is random in magnitude and direction for individual molecules, but a gas of many molecules has a predictable distribution of molecular speeds. This distribution is called the *Maxwell-Boltzmann distribution*, after its originators, who calculated it based on kinetic theory, and has since been confirmed experimentally. (See [\[link\]](#).) The distribution has a long tail, because a few molecules may go several times the rms speed. The most probable speed v_p is less than the rms speed v_{rms} . [\[link\]](#) shows that the curve is shifted to higher speeds at higher temperatures, with a broader range of speeds.



The Maxwell-Boltzmann distribution of molecular speeds in an ideal gas. The most likely speed v_p is less than the rms speed v_{rms} . Although very high speeds are possible, only a tiny fraction of the molecules have speeds that are an order of magnitude greater than v_{rms} .

The distribution of thermal speeds depends strongly on temperature. As temperature increases, the speeds are shifted to higher values and the distribution is broadened.



The Maxwell-Boltzmann distribution is shifted to

higher speeds and is broadened at higher temperatures.

What is the implication of the change in distribution with temperature shown in [\[link\]](#) for humans? All other things being equal, if a person has a fever, he or she is likely to lose more water molecules, particularly from linings along moist cavities such as the lungs and mouth, creating a dry sensation in the mouth.

Example:

Calculating Temperature: Escape Velocity of Helium Atoms

In order to escape Earth's gravity, an object near the top of the atmosphere (at an altitude of 100 km) must travel away from Earth at 11.1 km/s. This speed is called the *escape velocity*. At what temperature would helium atoms have an rms speed equal to the escape velocity?

Strategy

Identify the knowns and unknowns and determine which equations to use to solve the problem.

Solution

1. Identify the knowns: v is the escape velocity, 11.1 km/s.
2. Identify the unknowns: We need to solve for temperature, T . We also need to solve for the mass m of the helium atom.
3. Determine which equations are needed.

- To solve for mass m of the helium atom, we can use information from the periodic table:

Equation:

$$m = \frac{\text{molar mass}}{\text{number of atoms per mole}}.$$

- To solve for temperature T , we can rearrange either

Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

or

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}}$$

to yield

Equation:

$$T = \frac{m\overline{v^2}}{3k},$$

where k is the Boltzmann constant and m is the mass of a helium atom.

4. Plug the known values into the equations and solve for the unknowns.

Equation:

$$m = \frac{\text{molar mass}}{\text{number of atoms per mole}} = \frac{4.0026 \times 10^{-3} \text{ kg/mol}}{6.02 \times 10^{23} \text{ mol}} = 6.65 \times 10^{-27} \text{ kg}$$

Equation:

$$T = \frac{(6.65 \times 10^{-27} \text{ kg})(11.1 \times 10^3 \text{ m/s})^2}{3(1.38 \times 10^{-23} \text{ J/K})} = 1.98 \times 10^4 \text{ K}$$

Discussion

This temperature is much higher than atmospheric temperature, which is approximately 250 K (-25°C or -10°F) at high altitude. Very few helium atoms are left in the atmosphere, but there were many when the atmosphere was formed. The reason for the loss of helium atoms is that there are a small number of helium atoms with speeds higher than Earth's escape velocity even at normal temperatures. The speed of a helium atom changes from one instant to the next, so that at any instant, there is a small, but nonzero chance that the speed is greater than the escape speed and the molecule escapes from Earth's gravitational pull. Heavier molecules, such as oxygen, nitrogen, and water (very little of which reach a very high altitude), have smaller rms speeds, and so it is much less likely that any of them will have speeds greater than the escape velocity. In fact, so few have speeds above the escape velocity that billions of years are required to lose significant amounts of the atmosphere. [\[link\]](#) shows the impact of a lack of an atmosphere on the Moon. Because the gravitational pull of the Moon is much weaker, it has lost almost its

entire atmosphere. The comparison between Earth and the Moon is discussed in this chapter's Problems and Exercises.



This photograph of Apollo 17 Commander Eugene Cernan driving the lunar rover on the Moon in 1972 looks as though it was taken at night with a large spotlight. In fact, the light is coming from the Sun. Because the acceleration due to gravity on the Moon is so low (about $1/6$ that of Earth), the Moon's escape velocity is much smaller. As a result, gas molecules escape very easily from the Moon, leaving it with virtually no atmosphere. Even during the daytime, the sky is black because there is no gas to scatter sunlight. (credit: Harrison H. Schmitt/NASA)

Exercise:**Check Your Understanding****Problem:**

If you consider a very small object such as a grain of pollen, in a gas, then the number of atoms and molecules striking its surface would also be relatively small. Would the grain of pollen experience any fluctuations in pressure due to statistical fluctuations in the number of gas atoms and molecules striking it in a given amount of time?

Solution:

Yes. Such fluctuations actually occur for a body of any size in a gas, but since the numbers of atoms and molecules are immense for macroscopic bodies, the fluctuations are a tiny percentage of the number of collisions, and the averages spoken of in this section vary imperceptibly. Roughly speaking the fluctuations are proportional to the inverse square root of the number of collisions, so for small bodies they can become significant. This was actually observed in the 19th century for pollen grains in water, and is known as the Brownian effect.

Note:

PhET Explorations: Gas Properties

Pump gas molecules into a box and see what happens as you change the volume, add or remove heat, change gravity, and more. Measure the temperature and pressure, and discover how the properties of the gas vary in relation to each other.

[Gas](#)
[Propertie](#)
[s](#)

Section Summary

- Kinetic theory is the atomistic description of gases as well as liquids and solids.

- Kinetic theory models the properties of matter in terms of continuous random motion of atoms and molecules.
- The ideal gas law can also be expressed as

Equation:

$$PV = \frac{1}{3}Nm\overline{v^2},$$

where P is the pressure (average force per unit area), V is the volume of gas in the container, N is the number of molecules in the container, m is the mass of a molecule, and $\overline{v^2}$ is the average of the molecular speed squared.

- Thermal energy is defined to be the average translational kinetic energy $\overline{\text{KE}}$ of an atom or molecule.
- The temperature of gases is proportional to the average translational kinetic energy of atoms and molecules.

Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

or

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}}.$$

- The motion of individual molecules in a gas is random in magnitude and direction. However, a gas of many molecules has a predictable distribution of molecular speeds, known as the *Maxwell-Boltzmann distribution*.

Conceptual Questions

Exercise:

Problem:

How is momentum related to the pressure exerted by a gas? Explain on the atomic and molecular level, considering the behavior of atoms and molecules.

Problems & Exercises

Exercise:**Problem:**

Some incandescent light bulbs are filled with argon gas. What is v_{rms} for argon atoms near the filament, assuming their temperature is 2500 K?

Solution:

$$1.25 \times 10^3 \text{ m/s}$$

Exercise:**Problem:**

Average atomic and molecular speeds (v_{rms}) are large, even at low temperatures. What is v_{rms} for helium atoms at 5.00 K, just one degree above helium's liquefaction temperature?

Exercise:**Problem:**

(a) What is the average kinetic energy in joules of hydrogen atoms on the 5500°C surface of the Sun? (b) What is the average kinetic energy of helium atoms in a region of the solar corona where the temperature is $6.00 \times 10^5 \text{ K}$?

Solution:

(a) $1.20 \times 10^{-19} \text{ J}$

(b) $1.24 \times 10^{-17} \text{ J}$

Exercise:**Problem:**

The escape velocity of any object from Earth is 11.2 km/s. (a) Express this speed in m/s and km/h. (b) At what temperature would oxygen molecules (molecular mass is equal to 32.0 g/mol) have an average velocity v_{rms} equal to Earth's escape velocity of 11.1 km/s?

Exercise:

Problem:

The escape velocity from the Moon is much smaller than from Earth and is only 2.38 km/s. At what temperature would hydrogen molecules (molecular mass is equal to 2.016 g/mol) have an average velocity v_{rms} equal to the Moon's escape velocity?

Solution:

458 K

Exercise:**Problem:**

Nuclear fusion, the energy source of the Sun, hydrogen bombs, and fusion reactors, occurs much more readily when the average kinetic energy of the atoms is high—that is, at high temperatures. Suppose you want the atoms in your fusion experiment to have average kinetic energies of 6.40×10^{-14} J. What temperature is needed?

Exercise:**Problem:**

Suppose that the average velocity (v_{rms}) of carbon dioxide molecules (molecular mass is equal to 44.0 g/mol) in a flame is found to be 1.05×10^5 m/s. What temperature does this represent?

Solution:

1.95×10^7 K

Exercise:**Problem:**

Hydrogen molecules (molecular mass is equal to 2.016 g/mol) have an average velocity v_{rms} equal to 193 m/s. What is the temperature?

Exercise:

Problem:

Much of the gas near the Sun is atomic hydrogen. Its temperature would have to be 1.5×10^7 K for the average velocity v_{rms} to equal the escape velocity from the Sun. What is that velocity?

Solution:

$$6.09 \times 10^5 \text{ m/s}$$

Exercise:**Problem:**

There are two important isotopes of uranium— ^{235}U and ^{238}U ; these isotopes are nearly identical chemically but have different atomic masses. Only ^{235}U is very useful in nuclear reactors. One of the techniques for separating them (gas diffusion) is based on the different average velocities v_{rms} of uranium hexafluoride gas, UF_6 . (a) The molecular masses for $^{235}\text{U UF}_6$ and $^{238}\text{U UF}_6$ are 349.0 g/mol and 352.0 g/mol, respectively. What is the ratio of their average velocities? (b) At what temperature would their average velocities differ by 1.00 m/s? (c) Do your answers in this problem imply that this technique may be difficult?

Glossary

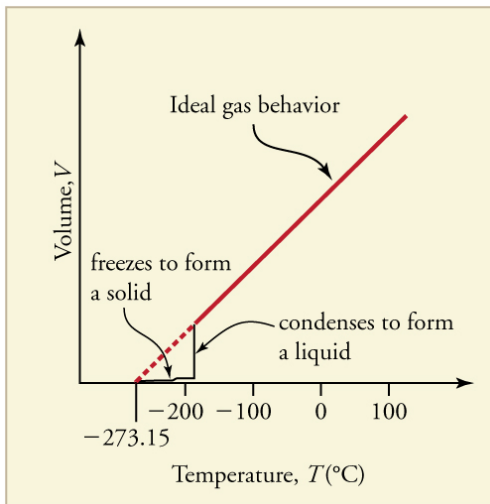
thermal energy

$\overline{\text{KE}}$, the average translational kinetic energy of a molecule

Phase Changes

- Interpret a phase diagram.
- State Dalton's law.
- Identify and describe the triple point of a gas from its phase diagram.
- Describe the state of equilibrium between a liquid and a gas, a liquid and a solid, and a gas and a solid.

Up to now, we have considered the behavior of ideal gases. Real gases are like ideal gases at high temperatures. At lower temperatures, however, the interactions between the molecules and their volumes cannot be ignored. The molecules are very close (condensation occurs) and there is a dramatic decrease in volume, as seen in [\[link\]](#). The substance changes from a gas to a liquid. When a liquid is cooled to even lower temperatures, it becomes a solid. The volume never reaches zero because of the finite volume of the molecules.



A sketch of volume versus temperature for a real gas at constant pressure. The linear (straight line) part of the graph represents ideal gas behavior—volume and temperature are directly and positively related and

the line extrapolates to zero volume at -273.15°C , or absolute zero. When the gas becomes a liquid, however, the volume actually decreases precipitously at the liquefaction point. The volume decreases slightly once the substance is solid, but it never becomes zero.

High pressure may also cause a gas to change phase to a liquid. Carbon dioxide, for example, is a gas at room temperature and atmospheric pressure, but becomes a liquid under sufficiently high pressure. If the pressure is reduced, the temperature drops and the liquid carbon dioxide solidifies into a snow-like substance at the temperature -78°C . Solid CO_2 is called “dry ice.” Another example of a gas that can be in a liquid phase is liquid nitrogen (LN_2). LN_2 is made by liquefaction of atmospheric air (through compression and cooling). It boils at 77 K (-196°C) at atmospheric pressure. LN_2 is useful as a refrigerant and allows for the preservation of blood, sperm, and other biological materials. It is also used to reduce noise in electronic sensors and equipment, and to help cool down their current-carrying wires. In dermatology, LN_2 is used to freeze and painlessly remove warts and other growths from the skin.

PV Diagrams

We can examine aspects of the behavior of a substance by plotting a graph of pressure versus volume, called a **PV diagram**. When the substance behaves like an ideal gas, the ideal gas law describes the relationship between its pressure and volume. That is,

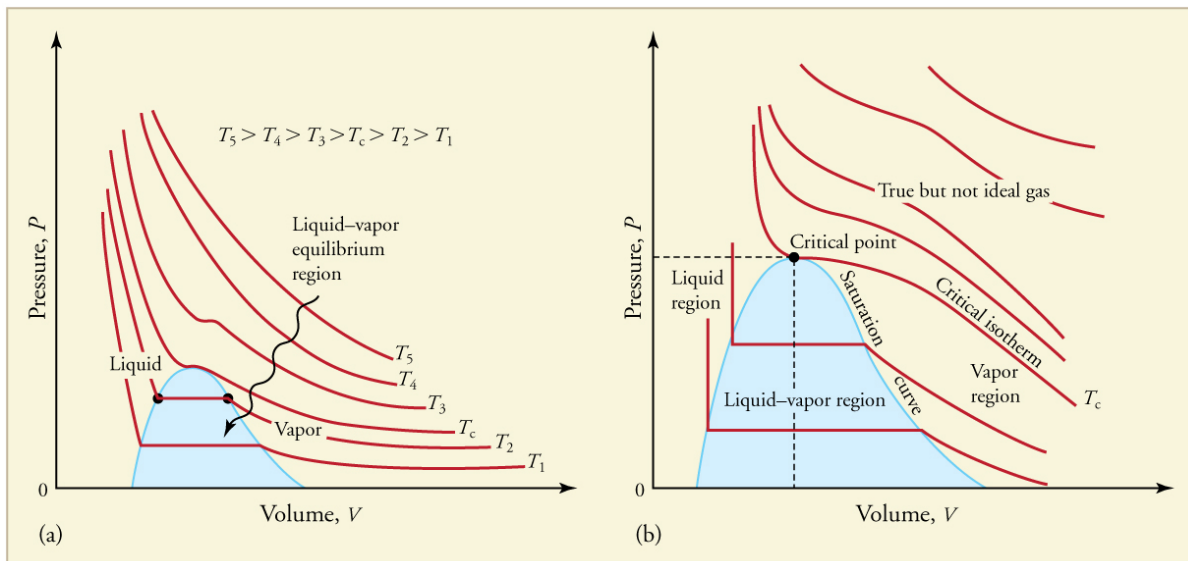
Equation:

$$PV = NkT \text{ (ideal gas).}$$

Now, assuming the number of molecules and the temperature are fixed,
Equation:

$$PV = \text{constant (ideal gas, constant temperature).}$$

For example, the volume of the gas will decrease as the pressure increases. If you plot the relationship $PV = \text{constant}$ on a PV diagram, you find a hyperbola. [\[link\]](#) shows a graph of pressure versus volume. The hyperbolas represent ideal-gas behavior at various fixed temperatures, and are called *isotherms*. At lower temperatures, the curves begin to look less like hyperbolas—the gas is not behaving ideally and may even contain liquid. There is a **critical point**—that is, a **critical temperature**—above which liquid cannot exist. At sufficiently high pressure above the critical point, the gas will have the density of a liquid but will not condense. Carbon dioxide, for example, cannot be liquefied at a temperature above 31.0°C. **Critical pressure** is the minimum pressure needed for liquid to exist at the critical temperature. [\[link\]](#) lists representative critical temperatures and pressures.



PV diagrams. (a) Each curve (isotherm) represents the relationship

between P and V at a fixed temperature; the upper curves are at higher temperatures. The lower curves are not hyperbolas, because the gas is no longer an ideal gas. (b) An expanded portion of the PV diagram for low temperatures, where the phase can change from a gas to a liquid. The term “vapor” refers to the gas phase when it exists at a temperature below the boiling temperature.

Substance	Critical temperature		Critical pressure	
	K	°C	Pa	atm
Water	647.4	374.3	22.12×10^6	219.0
Sulfur dioxide	430.7	157.6	7.88×10^6	78.0
Ammonia	405.5	132.4	11.28×10^6	111.7
Carbon dioxide	304.2	31.1	7.39×10^6	73.2

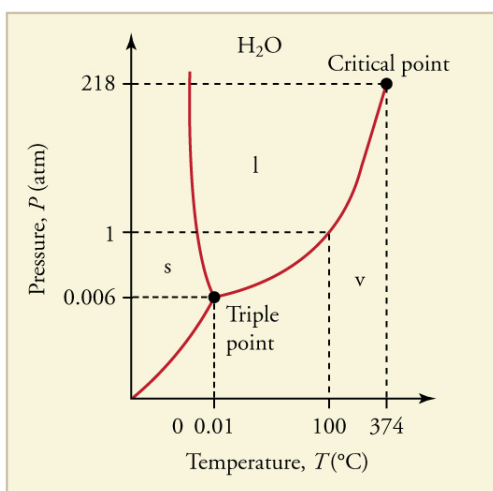
Substance	Critical temperature		Critical pressure	
	K	°C	Pa	atm
Oxygen	154.8	−118.4	5.08×10^6	50.3
Nitrogen	126.2	−146.9	3.39×10^6	33.6
Hydrogen	33.3	−239.9	1.30×10^6	12.9
Helium	5.3	−267.9	0.229×10^6	2.27

Critical Temperatures and Pressures

Phase Diagrams

The plots of pressure versus temperatures provide considerable insight into thermal properties of substances. There are well-defined regions on these graphs that correspond to various phases of matter, so PT graphs are called **phase diagrams**. [\[link\]](#) shows the phase diagram for water. Using the graph, if you know the pressure and temperature you can determine the phase of water. The solid lines—boundaries between phases—indicate temperatures and pressures at which the phases coexist (that is, they exist together in ratios, depending on pressure and temperature). For example, the boiling point of water is 100°C at 1.00 atm. As the pressure increases, the boiling temperature rises steadily to 374°C at a pressure of 218 atm. A pressure cooker (or even a covered pot) will cook food faster because the

water can exist as a liquid at temperatures greater than 100°C without all boiling away. The curve ends at a point called the *critical point*, because at higher temperatures the liquid phase does not exist at any pressure. The critical point occurs at the critical temperature, as you can see for water from [\[link\]](#). The critical temperature for oxygen is -118°C , so oxygen cannot be liquefied above this temperature.



The phase diagram (PT graph) for water. Note that the axes are nonlinear and the graph is not to scale. This graph is simplified—there are several other exotic phases of ice at higher pressures.

Similarly, the curve between the solid and liquid regions in [\[link\]](#) gives the melting temperature at various pressures. For example, the melting point is 0°C at 1.00 atm, as expected. Note that, at a fixed temperature, you can change the phase from solid (ice) to liquid (water) by increasing the pressure. Ice melts from pressure in the hands of a snowball maker. From

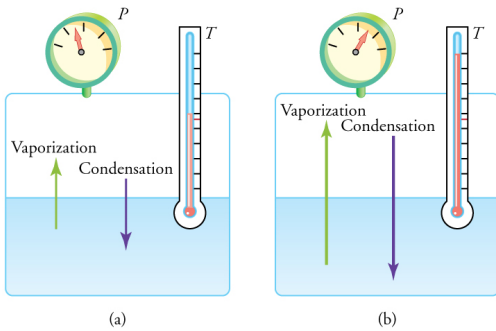
the phase diagram, we can also say that the melting temperature of ice rises with increased pressure. When a car is driven over snow, the increased pressure from the tires melts the snowflakes; afterwards the water refreezes and forms an ice layer.

At sufficiently low pressures there is no liquid phase, but the substance can exist as either gas or solid. For water, there is no liquid phase at pressures below 0.00600 atm. The phase change from solid to gas is called **sublimation**. It accounts for large losses of snow pack that never make it into a river, the routine automatic defrosting of a freezer, and the freeze-drying process applied to many foods. Carbon dioxide, on the other hand, sublimates at standard atmospheric pressure of 1 atm. (The solid form of CO_2 is known as dry ice because it does not melt. Instead, it moves directly from the solid to the gas state.)

All three curves on the phase diagram meet at a single point, the **triple point**, where all three phases exist in equilibrium. For water, the triple point occurs at 273.16 K (0.01°C), and is a more accurate calibration temperature than the melting point of water at 1.00 atm, or 273.15 K (0.0°C). See [\[link\]](#) for the triple point values of other substances.

Equilibrium

Liquid and gas phases are in equilibrium at the boiling temperature. (See [\[link\]](#).) If a substance is in a closed container at the boiling point, then the liquid is boiling and the gas is condensing at the same rate without net change in their relative amount. Molecules in the liquid escape as a gas at the same rate at which gas molecules stick to the liquid, or form droplets and become part of the liquid phase. The combination of temperature and pressure has to be “just right”; if the temperature and pressure are increased, equilibrium is maintained by the same increase of boiling and condensation rates.



Equilibrium between liquid and gas at two different boiling points inside a closed container. (a) The rates of boiling and condensation are equal at this combination of temperature and pressure, so the liquid and gas phases are in equilibrium. (b) At a higher temperature, the boiling rate is faster and the rates at which molecules leave the liquid and enter the gas are also faster. Because there are more molecules in the gas, the gas pressure is higher and the rate at which gas molecules condense and enter the liquid is faster. As a result the gas and liquid are in equilibrium at this higher temperature.

Substance	Temperature		Pressure	
	K	°C	Pa	atm
Water	273.16	0.01	6.10×10^2	0.00600
Carbon dioxide	216.55	-56.60	5.16×10^5	5.11
Sulfur dioxide	197.68	-75.47	1.67×10^3	0.0167
Ammonia	195.40	-77.75	6.06×10^3	0.0600
Nitrogen	63.18	-210.0	1.25×10^4	0.124
Oxygen	54.36	-218.8	1.52×10^2	0.00151
Hydrogen	13.84	-259.3	7.04×10^3	0.0697

Triple Point Temperatures and Pressures

One example of equilibrium between liquid and gas is that of water and steam at 100°C and 1.00 atm. This temperature is the boiling point at that pressure, so they should exist in equilibrium. Why does an open pot of water at 100°C boil completely away? The gas surrounding an open pot is

not pure water: it is mixed with air. If pure water and steam are in a closed container at 100°C and 1.00 atm , they would coexist—but with air over the pot, there are fewer water molecules to condense, and water boils. What about water at 20.0°C and 1.00 atm ? This temperature and pressure correspond to the liquid region, yet an open glass of water at this temperature will completely evaporate. Again, the gas around it is air and not pure water vapor, so that the reduced evaporation rate is greater than the condensation rate of water from dry air. If the glass is sealed, then the liquid phase remains. We call the gas phase a **vapor** when it exists, as it does for water at 20.0°C , at a temperature below the boiling temperature.

Exercise:

Check Your Understanding

Problem:

Explain why a cup of water (or soda) with ice cubes stays at 0°C , even on a hot summer day.

Solution:

The ice and liquid water are in thermal equilibrium, so that the temperature stays at the freezing temperature as long as ice remains in the liquid. (Once all of the ice melts, the water temperature will start to rise.)

Vapor Pressure, Partial Pressure, and Dalton's Law

Vapor pressure is defined as the pressure at which a gas coexists with its solid or liquid phase. Vapor pressure is created by faster molecules that break away from the liquid or solid and enter the gas phase. The vapor pressure of a substance depends on both the substance and its temperature—an increase in temperature increases the vapor pressure.

Partial pressure is defined as the pressure a gas would create if it occupied the total volume available. In a mixture of gases, *the total pressure is the sum of partial pressures of the component gases*, assuming ideal gas behavior and no chemical reactions between the components. This law is

known as **Dalton's law of partial pressures**, after the English scientist John Dalton (1766–1844), who proposed it. Dalton's law is based on kinetic theory, where each gas creates its pressure by molecular collisions, independent of other gases present. It is consistent with the fact that pressures add according to [Pascal's Principle](#). Thus water evaporates and ice sublimates when their vapor pressures exceed the partial pressure of water vapor in the surrounding mixture of gases. If their vapor pressures are less than the partial pressure of water vapor in the surrounding gas, liquid droplets or ice crystals (frost) form.

Exercise:

Check Your Understanding

Problem:

Is energy transfer involved in a phase change? If so, will energy have to be supplied to change phase from solid to liquid and liquid to gas? What about gas to liquid and liquid to solid? Why do they spray the orange trees with water in Florida when the temperatures are near or just below freezing?

Solution:

Yes, energy transfer is involved in a phase change. We know that atoms and molecules in solids and liquids are bound to each other because we know that force is required to separate them. So in a phase change from solid to liquid and liquid to gas, a force must be exerted, perhaps by collision, to separate atoms and molecules. Force exerted through a distance is work, and energy is needed to do work to go from solid to liquid and liquid to gas. This is intuitively consistent with the need for energy to melt ice or boil water. The converse is also true. Going from gas to liquid or liquid to solid involves atoms and molecules pushing together, doing work and releasing energy.

Note:

PhET Explorations: States of Matter—Basics

Heat, cool, and compress atoms and molecules and watch as they change between solid, liquid, and gas phases.

https://phet.colorado.edu/sims/html/states-of-matter-basics/latest/states-of-matter-basics_en.html

Section Summary

- Most substances have three distinct phases: gas, liquid, and solid.
- Phase changes among the various phases of matter depend on temperature and pressure.
- The existence of the three phases with respect to pressure and temperature can be described in a phase diagram.
- Two phases coexist (i.e., they are in thermal equilibrium) at a set of pressures and temperatures. These are described as a line on a phase diagram.
- The three phases coexist at a single pressure and temperature. This is known as the triple point and is described by a single point on a phase diagram.
- A gas at a temperature below its boiling point is called a vapor.
- Vapor pressure is the pressure at which a gas coexists with its solid or liquid phase.
- Partial pressure is the pressure a gas would create if it existed alone.
- Dalton's law states that the total pressure is the sum of the partial pressures of all of the gases present.

Conceptual Questions

Exercise:

Problem:

A pressure cooker contains water and steam in equilibrium at a pressure greater than atmospheric pressure. How does this greater pressure increase cooking speed?

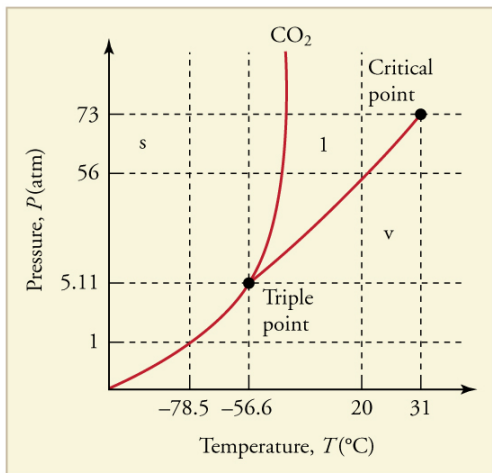
Exercise:

Problem:

Why does condensation form most rapidly on the coldest object in a room—for example, on a glass of ice water?

Exercise:**Problem:**

What is the vapor pressure of solid carbon dioxide (dry ice) at -78.5°C ?



The phase diagram for carbon dioxide. The axes are nonlinear, and the graph is not to scale. Dry ice is solid carbon dioxide and has a sublimation temperature of -78.5°C .

Exercise:**Problem:**

Can carbon dioxide be liquefied at room temperature (20°C)? If so, how? If not, why not? (See [\[link\]](#).)

Exercise:**Problem:**

Oxygen cannot be liquefied at room temperature by placing it under a large enough pressure to force its molecules together. Explain why this is.

Exercise:

Problem: What is the distinction between gas and vapor?

Glossary

PV diagram

a graph of pressure vs. volume

critical point

the temperature above which a liquid cannot exist

critical temperature

the temperature above which a liquid cannot exist

critical pressure

the minimum pressure needed for a liquid to exist at the critical temperature

vapor

a gas at a temperature below the boiling temperature

vapor pressure

the pressure at which a gas coexists with its solid or liquid phase

phase diagram

a graph of pressure vs. temperature of a particular substance, showing at which pressures and temperatures the three phases of the substance occur

triple point

the pressure and temperature at which a substance exists in equilibrium as a solid, liquid, and gas

sublimation

the phase change from solid to gas

partial pressure

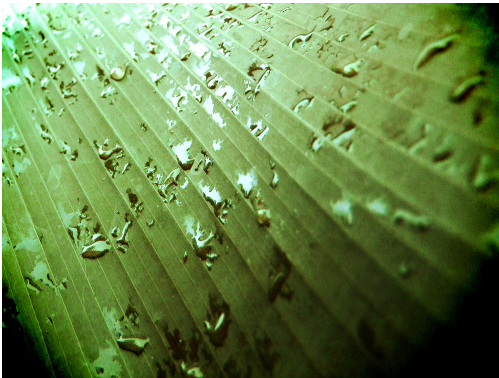
the pressure a gas would create if it occupied the total volume of space available

Dalton's law of partial pressures

the physical law that states that the total pressure of a gas is the sum of partial pressures of the component gases

Humidity, Evaporation, and Boiling

- Explain the relationship between vapor pressure of water and the capacity of air to hold water vapor.
- Explain the relationship between relative humidity and partial pressure of water vapor in the air.
- Calculate vapor density using vapor pressure.
- Calculate humidity and dew point.



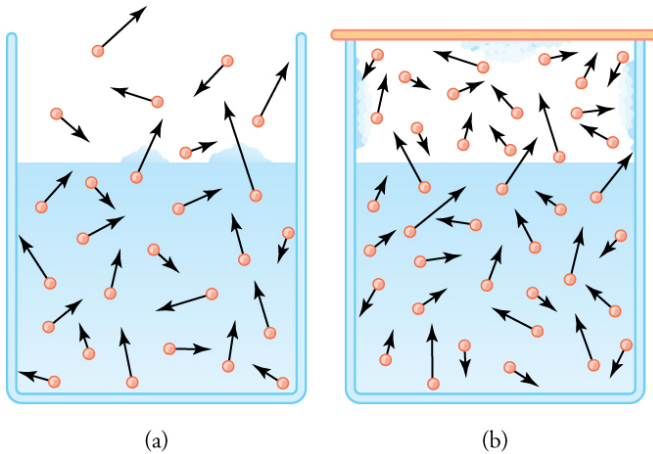
Dew drops like these, on a banana leaf photographed just after sunrise, form when the air temperature drops to or below the dew point. At the dew point, the rate at which water molecules join together is greater than the rate at which they separate, and some of the water condenses to form droplets. (credit: Aaron Escobar, Flickr)

The expression “it’s not the heat, it’s the humidity” makes a valid point. We keep cool in hot weather by evaporating sweat from our skin and water

from our breathing passages. Because evaporation is inhibited by high humidity, we feel hotter at a given temperature when the humidity is high. Low humidity, on the other hand, can cause discomfort from excessive drying of mucous membranes and can lead to an increased risk of respiratory infections.

When we say humidity, we really mean **relative humidity**. Relative humidity tells us how much water vapor is in the air compared with the maximum possible. At its maximum, denoted as **saturation**, the relative humidity is 100%, and evaporation is inhibited. The amount of water vapor in the air depends on temperature. For example, relative humidity rises in the evening, as air temperature declines, sometimes reaching the **dew point**. At the dew point temperature, relative humidity is 100%, and fog may result from the condensation of water droplets if they are small enough to stay in suspension. Conversely, if you wish to dry something (perhaps your hair), it is more effective to blow hot air over it rather than cold air, because, among other things, the increase in temperature increases the energy of the molecules, so the rate of evaporation increases.

The amount of water vapor in the air depends on the vapor pressure of water. The liquid and solid phases are continuously giving off vapor because some of the molecules have high enough speeds to enter the gas phase; see [\[link\]](#)(a). If a lid is placed over the container, as in [\[link\]](#)(b), evaporation continues, increasing the pressure, until sufficient vapor has built up for condensation to balance evaporation. Then equilibrium has been achieved, and the vapor pressure is equal to the partial pressure of water in the container. Vapor pressure increases with temperature because molecular speeds are higher as temperature increases. [\[link\]](#) gives representative values of water vapor pressure over a range of temperatures.



(a) Because of the distribution of speeds and kinetic energies, some water molecules can break away to the vapor phase even at temperatures below the ordinary boiling point. (b) If the container is sealed, evaporation will continue until there is enough vapor density for the condensation rate to equal the evaporation rate. This vapor density and the partial pressure it creates are the saturation values. They increase with temperature and are independent of the presence of other gases, such as air. They depend only on the vapor pressure of water.

Relative humidity is related to the partial pressure of water vapor in the air. At 100% humidity, the partial pressure is equal to the vapor pressure, and no more water can enter the vapor phase. If the partial pressure is less than the vapor pressure, then evaporation will take place, as humidity is less than 100%. If the partial pressure is greater than the vapor pressure, condensation takes place. In everyday language, people sometimes refer to

the capacity of air to “hold” water vapor, but this is not actually what happens. The water vapor is not held by the air. The amount of water in air is determined by the vapor pressure of water and has nothing to do with the properties of air.

Temperature (°C)	Vapor pressure (Pa)	Saturation vapor density (g/m ³)
−50	4.0	0.039
−20	1.04×10^2	0.89
−10	2.60×10^2	2.36
0	6.10×10^2	4.84
5	8.68×10^2	6.80
10	1.19×10^3	9.40

Temperature (°C)	Vapor pressure (Pa)	Saturation vapor density (g/m ³)
15	1.69×10^3	12.8
20	2.33×10^3	17.2
25	3.17×10^3	23.0
30	4.24×10^3	30.4
37	6.31×10^3	44.0
40	7.34×10^3	51.1
50	1.23×10^4	82.4
60	1.99×10^4	130
70	3.12×10^4	197

Temperature (°C)	Vapor pressure (Pa)	Saturation vapor density (g/m ³)
80	4.73×10^4	294
90	7.01×10^4	418
95	8.59×10^4	505
100	1.01×10^5	598
120	1.99×10^5	1095
150	4.76×10^5	2430
200	1.55×10^6	7090
220	2.32×10^6	10,200

Saturation Vapor Density of Water

Example:**Calculating Density Using Vapor Pressure**

[\[link\]](#) gives the vapor pressure of water at 20.0°C as 2.33×10^3 Pa. Use the ideal gas law to calculate the density of water vapor in g/m³ that would create a partial pressure equal to this vapor pressure. Compare the result with the saturation vapor density given in the table.

Strategy

To solve this problem, we need to break it down into a two steps. The partial pressure follows the ideal gas law,

Equation:

$$PV = nRT,$$

where n is the number of moles. If we solve this equation for n/V to calculate the number of moles per cubic meter, we can then convert this quantity to grams per cubic meter as requested. To do this, we need to use the molecular mass of water, which is given in the periodic table.

Solution

1. Identify the knowns and convert them to the proper units:

- a. temperature $T = 20^\circ\text{C} = 293$ K
- b. vapor pressure P of water at 20°C is 2.33×10^3 Pa
- c. molecular mass of water is 18.0 g/mol

2. Solve the ideal gas law for n/V .

Equation:

$$\frac{n}{V} = \frac{P}{RT}$$

3. Substitute known values into the equation and solve for n/V .

Equation:

$$\frac{n}{V} = \frac{P}{RT} = \frac{2.33 \times 10^3 \text{ Pa}}{(8.31 \text{ J/mol} \cdot \text{K})(293 \text{ K})} = 0.957 \text{ mol/m}^3$$

4. Convert the density in moles per cubic meter to grams per cubic meter.

Equation:

$$\rho = \left(0.957 \frac{\text{mol}}{\text{m}^3}\right) \left(\frac{18.0 \text{ g}}{\text{mol}}\right) = 17.2 \text{ g/m}^3$$

Discussion

The density is obtained by assuming a pressure equal to the vapor pressure of water at 20.0°C. The density found is identical to the value in [\[link\]](#), which means that a vapor density of 17.2 g/m³ at 20.0°C creates a partial pressure of 2.33×10^3 Pa, equal to the vapor pressure of water at that temperature. If the partial pressure is equal to the vapor pressure, then the liquid and vapor phases are in equilibrium, and the relative humidity is 100%. Thus, there can be no more than 17.2 g of water vapor per m³ at 20.0°C, so that this value is the saturation vapor density at that temperature. This example illustrates how water vapor behaves like an ideal gas: the pressure and density are consistent with the ideal gas law (assuming the density in the table is correct). The saturation vapor densities listed in [\[link\]](#) are the maximum amounts of water vapor that air can hold at various temperatures.

Note:**Percent Relative Humidity**

We define **percent relative humidity** as the ratio of vapor density to saturation vapor density, or

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100$$

We can use this and the data in [\[link\]](#) to do a variety of interesting calculations, keeping in mind that relative humidity is based on the comparison of the partial pressure of water vapor in air and ice.

Example:**Calculating Humidity and Dew Point**

(a) Calculate the percent relative humidity on a day when the temperature is 25.0°C and the air contains 9.40 g of water vapor per m³. (b) At what temperature will this air reach 100% relative humidity (the saturation density)? This temperature is the dew point. (c) What is the humidity when the air temperature is 25.0°C and the dew point is – 10.0°C?

Strategy and Solution

(a) Percent relative humidity is defined as the ratio of vapor density to saturation vapor density.

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100$$

The first is given to be 9.40 g/m³, and the second is found in [\[link\]](#) to be 23.0 g/m³. Thus,

Equation:

$$\text{percent relative humidity} = \frac{9.40 \text{ g/m}^3}{23.0 \text{ g/m}^3} \times 100 = 40.9\%$$

(b) The air contains 9.40 g/m³ of water vapor. The relative humidity will be 100% at a temperature where 9.40 g/m³ is the saturation density. Inspection of [\[link\]](#) reveals this to be the case at 10.0°C, where the relative humidity will be 100%. That temperature is called the dew point for air with this concentration of water vapor.

(c) Here, the dew point temperature is given to be – 10.0°C. Using [\[link\]](#), we see that the vapor density is 2.36 g/m³, because this value is the saturation vapor density at – 10.0°C. The saturation vapor density at 25.0°C is seen to be 23.0 g/m³. Thus, the relative humidity at 25.0°C is

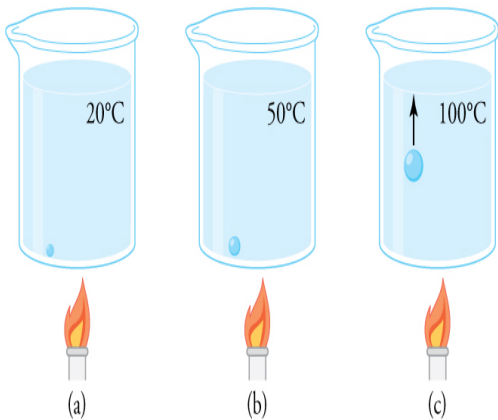
Equation:

$$\text{percent relative humidity} = \frac{2.36 \text{ g/m}^3}{23.0 \text{ g/m}^3} \times 100 = 10.3\%.$$

Discussion

The importance of dew point is that air temperature cannot drop below 10.0°C in part (b), or -10.0°C in part (c), without water vapor condensing out of the air. If condensation occurs, considerable transfer of heat occurs (discussed in [Heat and Heat Transfer Methods](#)), which prevents the temperature from further dropping. When dew points are below 0°C , freezing temperatures are a greater possibility, which explains why farmers keep track of the dew point. Low humidity in deserts means low dew-point temperatures. Thus condensation is unlikely. If the temperature drops, vapor does not condense in liquid drops. Because no heat is released into the air, the air temperature drops more rapidly compared to air with higher humidity. Likewise, at high temperatures, liquid droplets do not evaporate, so that no heat is removed from the gas to the liquid phase. This explains the large range of temperature in arid regions.

Why does water boil at 100°C ? You will note from [\[link\]](#) that the vapor pressure of water at 100°C is $1.01 \times 10^5 \text{ Pa}$, or 1.00 atm. Thus, it can evaporate without limit at this temperature and pressure. But why does it form bubbles when it boils? This is because water ordinarily contains significant amounts of dissolved air and other impurities, which are observed as small bubbles of air in a glass of water. If a bubble starts out at the bottom of the container at 20°C , it contains water vapor (about 2.30%). The pressure inside the bubble is fixed at 1.00 atm (we ignore the slight pressure exerted by the water around it). As the temperature rises, the amount of air in the bubble stays the same, but the water vapor increases; the bubble expands to keep the pressure at 1.00 atm. At 100°C , water vapor enters the bubble continuously since the partial pressure of water is equal to 1.00 atm in equilibrium. It cannot reach this pressure, however, since the bubble also contains air and total pressure is 1.00 atm. The bubble grows in size and thereby increases the buoyant force. The bubble breaks away and rises rapidly to the surface—we call this boiling! (See [\[link\]](#).)



- (a) An air bubble in water starts out saturated with water vapor at 20°C. (b) As the temperature rises, water vapor enters the bubble because its vapor pressure increases. The bubble expands to keep its pressure at 1.00 atm. (c) At 100°C, water vapor enters the bubble continuously because water's vapor pressure exceeds its partial pressure in the bubble, which must be less than 1.00 atm. The bubble grows and rises to the surface.

Exercise:
Check Your Understanding

Problem:

Freeze drying is a process in which substances, such as foods, are dried by placing them in a vacuum chamber and lowering the atmospheric pressure around them. How does the lowered atmospheric pressure speed the drying process, and why does it cause the temperature of the food to drop?

Solution:

Decreased the atmospheric pressure results in decreased partial pressure of water, hence a lower humidity. So evaporation of water from food, for example, will be enhanced. The molecules of water most likely to break away from the food will be those with the greatest velocities. Those remaining thus have a lower average velocity and a lower temperature. This can (and does) result in the freezing and drying of the food; hence the process is aptly named freeze drying.

Note:

PhET Explorations: States of Matter

Watch different types of molecules form a solid, liquid, or gas. Add or remove heat and watch the phase change. Change the temperature or volume of a container and see a pressure-temperature diagram respond in real time. Relate the interaction potential to the forces between molecules.

https://phet.colorado.edu/sims/html/states-of-matter/latest/states-of-matter_en.html

Section Summary

- Relative humidity is the fraction of water vapor in a gas compared to the saturation value.
- The saturation vapor density can be determined from the vapor pressure for a given temperature.

- Percent relative humidity is defined to be

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100.$$

- The dew point is the temperature at which air reaches 100% relative humidity.

Conceptual Questions

Exercise:

Problem:

Because humidity depends only on water's vapor pressure and temperature, are the saturation vapor densities listed in [\[link\]](#) valid in an atmosphere of helium at a pressure of $1.01 \times 10^5 \text{ N/m}^2$, rather than air? Are those values affected by altitude on Earth?

Exercise:

Problem:

Why does a beaker of 40.0°C water placed in a vacuum chamber start to boil as the chamber is evacuated (air is pumped out of the chamber)? At what pressure does the boiling begin? Would food cook any faster in such a beaker?

Exercise:

Problem:

Why does rubbing alcohol evaporate much more rapidly than water at STP (standard temperature and pressure)?

Problems & Exercises

Exercise:

Problem:

Dry air is 78.1% nitrogen. What is the partial pressure of nitrogen when the atmospheric pressure is $1.01 \times 10^5 \text{ N/m}^2$?

Solution:

$$7.89 \times 10^4 \text{ Pa}$$

Exercise:**Problem:**

(a) What is the vapor pressure of water at 20.0°C ? (b) What percentage of atmospheric pressure does this correspond to? (c) What percent of 20.0°C air is water vapor if it has 100% relative humidity? (The density of dry air at 20.0°C is 1.20 kg/m^3 .)

Exercise:**Problem:**

Pressure cookers increase cooking speed by raising the boiling temperature of water above its value at atmospheric pressure. (a) What pressure is necessary to raise the boiling point to 120.0°C ? (b) What gauge pressure does this correspond to?

Solution:

(a) $1.99 \times 10^5 \text{ Pa}$

(b) 0.97 atm

Exercise:

Problem:

(a) At what temperature does water boil at an altitude of 1500 m (about 5000 ft) on a day when atmospheric pressure is $8.59 \times 10^4 \text{ N/m}^2$? (b) What about at an altitude of 3000 m (about 10,000 ft) when atmospheric pressure is $7.00 \times 10^4 \text{ N/m}^2$?

Exercise:**Problem:**

What is the atmospheric pressure on top of Mt. Everest on a day when water boils there at a temperature of 70.0°C ?

Solution:

$$3.12 \times 10^4 \text{ Pa}$$

Exercise:**Problem:**

At a spot in the high Andes, water boils at 80.0°C , greatly reducing the cooking speed of potatoes, for example. What is atmospheric pressure at this location?

Exercise:**Problem:**

What is the relative humidity on a 25.0°C day when the air contains 18.0 g/m^3 of water vapor?

Solution:

$$78.3\%$$

Exercise:

Problem:

What is the density of water vapor in g/m^3 on a hot dry day in the desert when the temperature is 40.0°C and the relative humidity is 6.00%?

Exercise:**Problem:**

A deep-sea diver should breathe a gas mixture that has the same oxygen partial pressure as at sea level, where dry air contains 20.9% oxygen and has a total pressure of $1.01 \times 10^5 \text{ N}/\text{m}^2$. (a) What is the partial pressure of oxygen at sea level? (b) If the diver breathes a gas mixture at a pressure of $2.00 \times 10^6 \text{ N}/\text{m}^2$, what percent oxygen should it be to have the same oxygen partial pressure as at sea level?

Solution:

(a) $2.12 \times 10^4 \text{ Pa}$

(b) 1.06 %

Exercise:**Problem:**

The vapor pressure of water at 40.0°C is $7.34 \times 10^3 \text{ N}/\text{m}^2$. Using the ideal gas law, calculate the density of water vapor in g/m^3 that creates a partial pressure equal to this vapor pressure. The result should be the same as the saturation vapor density at that temperature ($51.1 \text{ g}/\text{m}^3$).

Exercise:

Problem:

Air in human lungs has a temperature of 37.0°C and a saturation vapor density of 44.0 g/m^3 . (a) If 2.00 L of air is exhaled and very dry air inhaled, what is the maximum loss of water vapor by the person? (b) Calculate the partial pressure of water vapor having this density, and compare it with the vapor pressure of $6.31 \times 10^3\text{ N/m}^2$.

Solution:

(a) $8.80 \times 10^{-2}\text{ g}$

(b) $6.30 \times 10^3\text{ Pa}$; the two values are nearly identical.

Exercise:**Problem:**

If the relative humidity is 90.0% on a muggy summer morning when the temperature is 20.0°C , what will it be later in the day when the temperature is 30.0°C , assuming the water vapor density remains constant?

Exercise:**Problem:**

Late on an autumn day, the relative humidity is 45.0% and the temperature is 20.0°C . What will the relative humidity be that evening when the temperature has dropped to 10.0°C , assuming constant water vapor density?

Solution:

82.3%

Exercise:

Problem:

Atmospheric pressure atop Mt. Everest is $3.30 \times 10^4 \text{ N/m}^2$. (a) What is the partial pressure of oxygen there if it is 20.9% of the air? (b) What percent oxygen should a mountain climber breathe so that its partial pressure is the same as at sea level, where atmospheric pressure is $1.01 \times 10^5 \text{ N/m}^2$? (c) One of the most severe problems for those climbing very high mountains is the extreme drying of breathing passages. Why does this drying occur?

Exercise:**Problem:**

What is the dew point (the temperature at which 100% relative humidity would occur) on a day when relative humidity is 39.0% at a temperature of 20.0°C ?

Solution:

4.77°C

Exercise:**Problem:**

On a certain day, the temperature is 25.0°C and the relative humidity is 90.0%. How many grams of water must condense out of each cubic meter of air if the temperature falls to 15.0°C ? Such a drop in temperature can, thus, produce heavy dew or fog.

Exercise:**Problem: Integrated Concepts**

The boiling point of water increases with depth because pressure increases with depth. At what depth will fresh water have a boiling point of 150°C , if the surface of the water is at sea level?

Solution:

38.3 m

Exercise:**Problem: Integrated Concepts**

(a) At what depth in fresh water is the critical pressure of water reached, given that the surface is at sea level? (b) At what temperature will this water boil? (c) Is a significantly higher temperature needed to boil water at a greater depth?

Exercise:**Problem: Integrated Concepts**

To get an idea of the small effect that temperature has on Archimedes' principle, calculate the fraction of a copper block's weight that is supported by the buoyant force in 0°C water and compare this fraction with the fraction supported in 95.0°C water.

Solution:

$\frac{(F_B/w_{Cu})}{(F_B/w_{Cu})'} = 1.02$. The buoyant force supports nearly the exact same amount of force on the copper block in both circumstances.

Exercise:**Problem: Integrated Concepts**

If you want to cook in water at 150°C, you need a pressure cooker that can withstand the necessary pressure. (a) What pressure is required for the boiling point of water to be this high? (b) If the lid of the pressure cooker is a disk 25.0 cm in diameter, what force must it be able to withstand at this pressure?

Exercise:

Problem: Unreasonable Results

(a) How many moles per cubic meter of an ideal gas are there at a pressure of $1.00 \times 10^{14} \text{ N/m}^2$ and at 0°C ? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) $4.41 \times 10^{10} \text{ mol/m}^3$

(b) It's unreasonably large.

(c) At high pressures such as these, the ideal gas law can no longer be applied. As a result, unreasonable answers come up when it is used.

Exercise:**Problem: Unreasonable Results**

(a) An automobile mechanic claims that an aluminum rod fits loosely into its hole on an aluminum engine block because the engine is hot and the rod is cold. If the hole is 10.0% bigger in diameter than the 22.0°C rod, at what temperature will the rod be the same size as the hole? (b) What is unreasonable about this temperature? (c) Which premise is responsible?

Exercise:**Problem: Unreasonable Results**

The temperature inside a supernova explosion is said to be $2.00 \times 10^{13} \text{ K}$. (a) What would the average velocity v_{rms} of hydrogen atoms be? (b) What is unreasonable about this velocity? (c) Which premise or assumption is responsible?

Solution:

(a) $7.03 \times 10^8 \text{ m/s}$

(b) The velocity is too high—it's greater than the speed of light.

(c) The assumption that hydrogen inside a supernova behaves as an idea gas is responsible, because of the great temperature and density in the core of a star. Furthermore, when a velocity greater than the speed of light is obtained, classical physics must be replaced by relativity, a subject not yet covered.

Exercise:

Problem: Unreasonable Results

Suppose the relative humidity is 80% on a day when the temperature is 30.0°C . (a) What will the relative humidity be if the air cools to 25.0°C and the vapor density remains constant? (b) What is unreasonable about this result? (c) Which premise is responsible?

Glossary

dew point

the temperature at which relative humidity is 100%; the temperature at which water starts to condense out of the air

saturation

the condition of 100% relative humidity

percent relative humidity

the ratio of vapor density to saturation vapor density

relative humidity

the amount of water in the air relative to the maximum amount the air can hold

Introduction to Heat and Heat Transfer Methods

class="introduction"

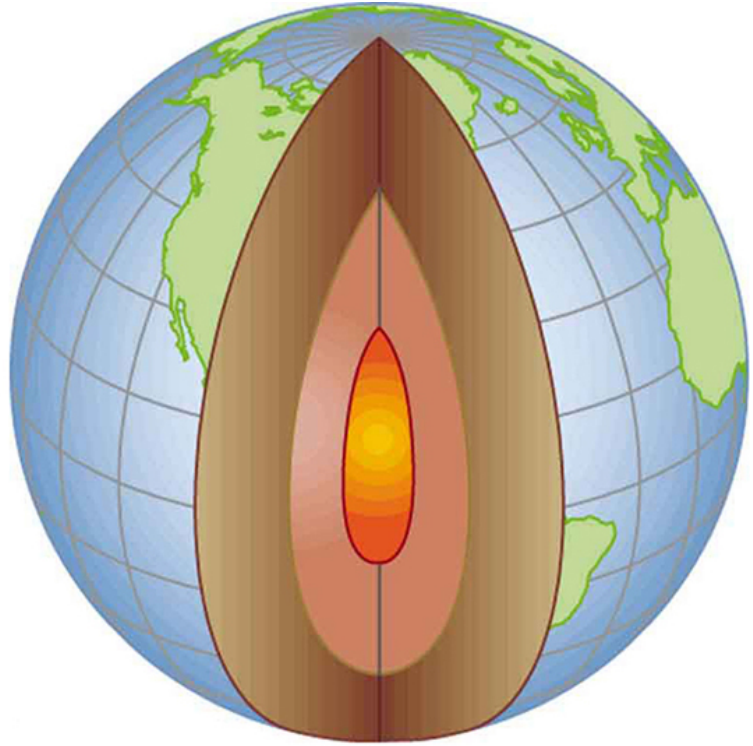
(a) The chilling effect of a clear breezy night is produced by the wind and by radiative heat transfer to cold outer space. (b) There was once great controversy about the Earth's age, but it is now generally accepted to be about 4.5 billion years old. Much of the debate is centered on the Earth's molten interior. According to our understanding of heat transfer, if the Earth is really that old, its

center should
have cooled
off long ago.

The
discovery of
radioactivity
in rocks
revealed the
source of
energy that
keeps the
Earth's
interior
molten,
despite heat
transfer to the
surface, and
from there to
cold outer
space.



(a)



(b)

Energy can exist in many forms and heat is one of the most intriguing. Heat is often hidden, as it only exists when in transit, and is transferred by a number of distinctly different methods. Heat transfer touches every aspect of our lives and helps us understand how the universe functions. It explains the chill we feel on a clear breezy night, or why Earth's core has yet to cool. This chapter defines and explores heat transfer, its effects, and the methods by which heat is transferred. These topics are fundamental, as well as practical, and will often be referred to in the chapters ahead.

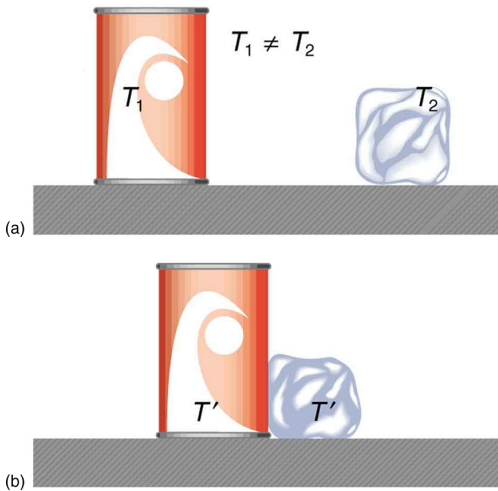
Heat

- Define heat as transfer of energy.

In [Work, Energy, and Energy Resources](#), we defined work as force times distance and learned that work done on an object changes its kinetic energy. We also saw in [Temperature, Kinetic Theory, and the Gas Laws](#) that temperature is proportional to the (average) kinetic energy of atoms and molecules. We say that a thermal system has a certain internal energy: its internal energy is higher if the temperature is higher. If two objects at different temperatures are brought in contact with each other, energy is transferred from the hotter to the colder object until equilibrium is reached and the bodies reach thermal equilibrium (i.e., they are at the same temperature). No work is done by either object, because no force acts through a distance. The transfer of energy is caused by the temperature difference, and ceases once the temperatures are equal. These observations lead to the following definition of **heat**: Heat is the spontaneous transfer of energy due to a temperature difference.

As noted in [Temperature, Kinetic Theory, and the Gas Laws](#), heat is often confused with temperature. For example, we may say the heat was unbearable, when we actually mean that the temperature was high. Heat is a form of energy, whereas temperature is not. The misconception arises because we are sensitive to the flow of heat, rather than the temperature.

Owing to the fact that heat is a form of energy, it has the SI unit of *joule* (J). The *calorie* (cal) is a common unit of energy, defined as the energy needed to change the temperature of 1.00 g of water by 1.00°C—specifically, between 14.5°C and 15.5°C, since there is a slight temperature dependence. Perhaps the most common unit of heat is the **kilocalorie** (kcal), which is the energy needed to change the temperature of 1.00 kg of water by 1.00°C. Since mass is most often specified in kilograms, kilocalorie is commonly used. Food calories (given the notation Cal, and sometimes called “big calorie”) are actually kilocalories (1 kilocalorie = 1000 calories), a fact not easily determined from package labeling.



In figure (a) the soft drink and the ice have different temperatures, T_1 and T_2 , and are not in thermal equilibrium. In figure (b), when the soft drink and ice are allowed to interact, energy is transferred until they reach the same temperature T' , achieving equilibrium. Heat transfer occurs due to the difference in temperatures. In fact, since the soft drink and ice are both in contact with the surrounding air and bench, the equilibrium temperature will be the same for both.

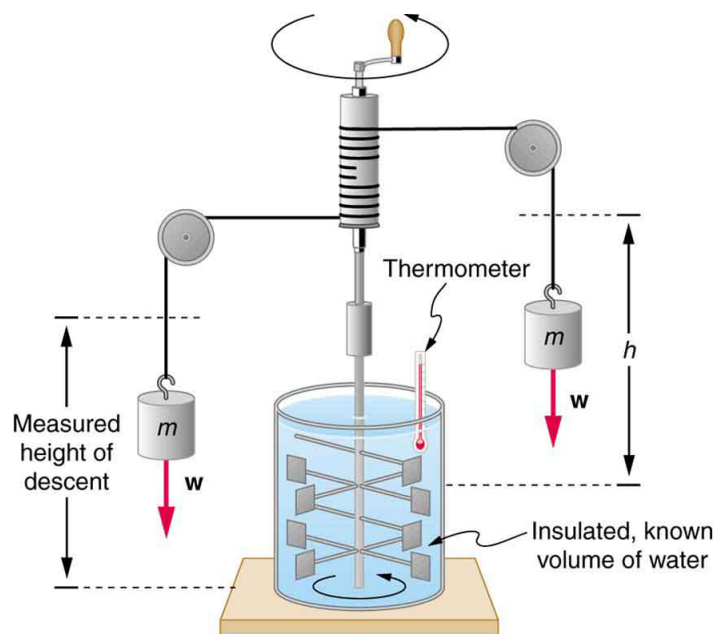
Mechanical Equivalent of Heat

It is also possible to change the temperature of a substance by doing work. Work can transfer energy into or out of a system. This realization helped establish the fact that heat is a form of energy. James Prescott Joule (1818–1889) performed many experiments to establish the **mechanical equivalent of heat**—*the work needed to produce the same effects as heat transfer*. In terms of the units used for these two terms, the best modern value for this equivalence is

Equation:

$$1.000 \text{ kcal} = 4186 \text{ J.}$$

We consider this equation as the conversion between two different units of energy.



Schematic depiction of Joule's experiment that established the equivalence of heat and work.

The figure above shows one of Joule's most famous experimental setups for demonstrating the mechanical equivalent of heat. It demonstrated that work and heat can produce the same effects, and helped establish the principle of conservation of energy. Gravitational potential energy (PE) (work done by the gravitational force) is converted into kinetic energy (KE), and then randomized by viscosity and turbulence into increased average kinetic energy of atoms and molecules in the system, producing a temperature increase. His contributions to the field of thermodynamics were so significant that the SI unit of energy was named after him.

Heat added or removed from a system changes its internal energy and thus its temperature. Such a temperature increase is observed while cooking. However, adding heat does not necessarily increase the temperature. An example is melting of ice; that is, when a substance changes from one phase to another. Work done on the system or by the system can also change the internal energy of the system. Joule demonstrated that the temperature of a system can be increased by stirring. If an ice cube is rubbed against a rough surface, work is done by the frictional force. A system has a well-defined internal energy, but we cannot say that it has a certain "heat content" or "work content". We use the phrase "heat transfer" to emphasize its nature.

Exercise:

Check Your Understanding

Problem:

Two samples (A and B) of the same substance are kept in a lab. Someone adds 10 kilojoules (kJ) of heat to one sample, while 10 kJ of work is done on the other sample. How can you tell to which sample the heat was added?

Solution:

Heat and work both change the internal energy of the substance. However, the properties of the sample only depend on the internal energy so that it is impossible to tell whether heat was added to sample A or B.

Summary

- Heat and work are the two distinct methods of energy transfer.
- Heat is energy transferred solely due to a temperature difference.
- Any energy unit can be used for heat transfer, and the most common are kilocalorie (kcal) and joule (J).
- Kilocalorie is defined to be the energy needed to change the temperature of 1.00 kg of water between 14.5°C and 15.5°C.
- The mechanical equivalent of this heat transfer is
 $1.00 \text{ kcal} = 4186 \text{ J}.$

Conceptual Questions

Exercise:

Problem: How is heat transfer related to temperature?

Exercise:

Problem:

Describe a situation in which heat transfer occurs. What are the resulting forms of energy?

Exercise:

Problem:

When heat transfers into a system, is the energy stored as heat? Explain briefly.

Glossary

heat

the spontaneous transfer of energy due to a temperature difference

kilocalorie

1 kilocalorie = 1000 calories

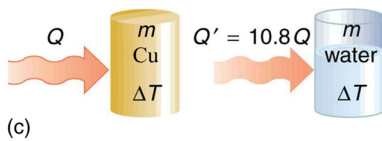
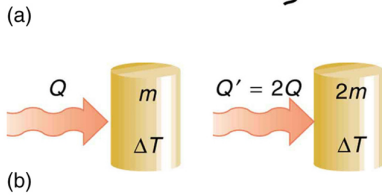
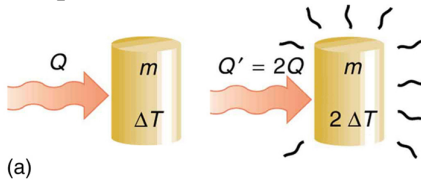
mechanical equivalent of heat

the work needed to produce the same effects as heat transfer

Temperature Change and Heat Capacity

- Observe heat transfer and change in temperature and mass.
- Calculate final temperature after heat transfer between two objects.

One of the major effects of heat transfer is temperature change: heating increases the temperature while cooling decreases it. We assume that there is no phase change and that no work is done on or by the system. Experiments show that the transferred heat depends on three factors—the change in temperature, the mass of the system, and the substance and phase of the substance.



The heat Q transferred to cause a temperature change depends on the magnitude of the temperature change, the mass of the system, and the substance and phase involved. (a) The amount of heat transferred is directly proportional to the temperature change.

To double the temperature change of a mass m , you need to add twice the heat. (b) The amount of heat transferred is also directly proportional to the mass.

To cause an equivalent temperature change in a

doubled mass, you need to add twice the heat. (c)

The amount of heat transferred depends on the substance and its phase. If it takes an amount Q of heat to cause a temperature change ΔT in a given mass of copper, it will take 10.8 times that amount of heat to cause the equivalent temperature change in the same mass of water assuming no phase change in either substance.

The dependence on temperature change and mass are easily understood. Owing to the fact that the (average) kinetic energy of an atom or molecule is proportional to the absolute temperature, the internal energy of a system is proportional to the absolute temperature and the number of atoms or molecules. Owing to the fact that the transferred heat is equal to the change in the internal energy, the heat is proportional to the mass of the substance and the temperature change. The transferred heat also depends on the substance so that, for example, the heat necessary to raise the temperature is less for alcohol than for water. For the same substance, the transferred heat also depends on the phase (gas, liquid, or solid).

Note:

Heat Transfer and Temperature Change

The quantitative relationship between heat transfer and temperature change contains all three factors:

Equation:

$$Q = mc\Delta T,$$

where Q is the symbol for heat transfer, m is the mass of the substance, and ΔT is the change in temperature. The symbol c stands for **specific heat** and depends on the material and phase. The specific heat is the amount of heat necessary to change the

temperature of 1.00 kg of mass by 1.00°C. The specific heat c is a property of the substance; its SI unit is J/(kg · K) or J/(kg · °C). Recall that the temperature change (ΔT) is the same in units of kelvin and degrees Celsius. If heat transfer is measured in kilocalories, then *the unit of specific heat* is kcal/(kg · °C).

Values of specific heat must generally be looked up in tables, because there is no simple way to calculate them. In general, the specific heat also depends on the temperature. [\[link\]](#) lists representative values of specific heat for various substances. Except for gases, the temperature and volume dependence of the specific heat of most substances is weak. We see from this table that the specific heat of water is five times that of glass and ten times that of iron, which means that it takes five times as much heat to raise the temperature of water the same amount as for glass and ten times as much heat to raise the temperature of water as for iron. In fact, water has one of the largest specific heats of any material, which is important for sustaining life on Earth.

Example:

Calculating the Required Heat: Heating Water in an Aluminum Pan

A 0.500 kg aluminum pan on a stove is used to heat 0.250 liters of water from 20.0°C to 80.0°C. (a) How much heat is required? What percentage of the heat is used to raise the temperature of (b) the pan and (c) the water?

Strategy

The pan and the water are always at the same temperature. When you put the pan on the stove, the temperature of the water and the pan is increased by the same amount. We use the equation for the heat transfer for the given temperature change and mass of water and aluminum. The specific heat values for water and aluminum are given in [\[link\]](#).

Solution

Because water is in thermal contact with the aluminum, the pan and the water are at the same temperature.

1. Calculate the temperature difference:

Equation:

$$\Delta T = T_f - T_i = 60.0^\circ\text{C}.$$

2. Calculate the mass of water. Because the density of water is 1000 kg/m³, one liter of water has a mass of 1 kg, and the mass of 0.250 liters of water is $m_w = 0.250$ kg.
3. Calculate the heat transferred to the water. Use the specific heat of water in [\[link\]](#):

Equation:

$$Q_w = m_w c_w \Delta T = (0.250 \text{ kg})(4186 \text{ J/kg}^\circ\text{C})(60.0^\circ\text{C}) = 62.8 \text{ kJ}.$$

4. Calculate the heat transferred to the aluminum. Use the specific heat for aluminum in [\[link\]](#):

Equation:

$$Q_{Al} = m_{Al} c_{Al} \Delta T = (0.500 \text{ kg})(900 \text{ J/kg}^\circ\text{C})(60.0^\circ\text{C}) = 27.0 \times 10^4 \text{ J} = 27.0 \text{ kJ}.$$

5. Compare the percentage of heat going into the pan versus that going into the water. First, find the total transferred heat:

Equation:

$$Q_{\text{Total}} = Q_W + Q_{Al} = 62.8 \text{ kJ} + 27.0 \text{ kJ} = 89.8 \text{ kJ}.$$

Thus, the amount of heat going into heating the pan is

Equation:

$$\frac{27.0 \text{ kJ}}{89.8 \text{ kJ}} \times 100\% = 30.1\%,$$

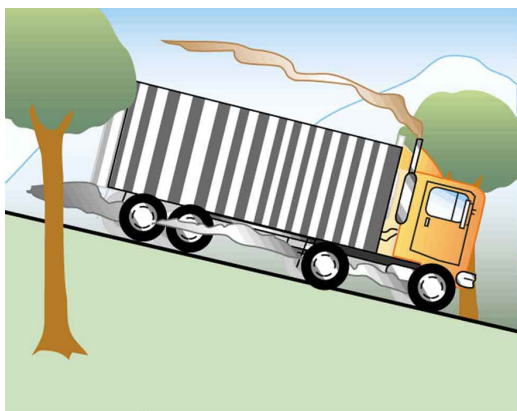
and the amount going into heating the water is

Equation:

$$\frac{62.8 \text{ kJ}}{89.8 \text{ kJ}} \times 100\% = 69.9\%.$$

Discussion

In this example, the heat transferred to the container is a significant fraction of the total transferred heat. Although the mass of the pan is twice that of the water, the specific heat of water is over four times greater than that of aluminum. Therefore, it takes a bit more than twice the heat to achieve the given temperature change for the water as compared to the aluminum pan.



The smoking brakes on this truck are a visible evidence of the mechanical equivalent of heat.

Example:

Calculating the Temperature Increase from the Work Done on a Substance: Truck Brakes Overheat on Downhill Runs

Truck brakes used to control speed on a downhill run do work, converting gravitational potential energy into increased internal energy (higher temperature) of the brake material. This conversion prevents the gravitational potential energy from being converted into kinetic energy of the truck. The problem is that the mass of the truck is large compared with that of the brake material absorbing the energy, and the temperature increase may occur too fast for sufficient heat to transfer from the brakes to the environment.

Calculate the temperature increase of 100 kg of brake material with an average specific heat of $800 \text{ J/kg} \cdot ^\circ\text{C}$ if the material retains 10% of the energy from a 10,000-kg truck descending 75.0 m (in vertical displacement) at a constant speed.

Strategy

If the brakes are not applied, gravitational potential energy is converted into kinetic energy. When brakes are applied, gravitational potential energy is converted into internal energy of the brake material. We first calculate the gravitational potential energy (Mgh) that the entire truck loses in its descent and then find the temperature increase produced in the brake material alone.

Solution

1. Calculate the change in gravitational potential energy as the truck goes downhill

Equation:

$$Mgh = (10,000 \text{ kg}) (9.80 \text{ m/s}^2) (75.0 \text{ m}) = 7.35 \times 10^6 \text{ J.}$$

2. Calculate the temperature from the heat transferred using $Q=Mgh$ and

Equation:

$$\Delta T = \frac{Q}{mc},$$

where m is the mass of the brake material. Insert the values $m = 100 \text{ kg}$ and $c = 800 \text{ J/kg} \cdot ^\circ\text{C}$ to find

Equation:

$$\Delta T = \frac{(7.35 \times 10^5 \text{ J})}{(100 \text{ kg})(800 \text{ J/kg}^\circ\text{C})} = 9.2^\circ\text{C.}$$

Discussion

This same idea underlies the recent hybrid technology of cars, where mechanical energy (gravitational potential energy) is converted by the brakes into electrical energy (battery).

Substances	Specific heat (c)	
Solids	J/kg·°C	kcal/kg·°C [footnote] These values are identical in units of cal/g ·°C.
Aluminum	900	0.215
Asbestos	800	0.19
Concrete, granite (average)	840	0.20
Copper	387	0.0924
Glass	840	0.20

Substances	Specific heat (c)	
Gold	129	0.0308
Human body (average at 37 °C)	3500	0.83
Ice (average, -50°C to 0°C)	2090	0.50
Iron, steel	452	0.108
Lead	128	0.0305
Silver	235	0.0562
Wood	1700	0.4
<i>Liquids</i>		
Benzene	1740	0.415
Ethanol	2450	0.586
Glycerin	2410	0.576
Mercury	139	0.0333
Water (15.0 °C)	4186	1.000
<i>Gases</i> [footnote] c_v at constant volume and at 20.0°C, except as noted, and at 1.00 atm average pressure. Values in parentheses are c_p at a constant pressure of 1.00 atm.		
Air (dry)	721 (1015)	0.172 (0.242)
Ammonia	1670 (2190)	0.399 (0.523)
Carbon dioxide	638 (833)	0.152 (0.199)

Substances	Specific heat (c)	
Nitrogen	739 (1040)	0.177 (0.248)
Oxygen	651 (913)	0.156 (0.218)
Steam (100°C)	1520 (2020)	0.363 (0.482)

Specific Heats^[footnote] of Various Substances

The values for solids and liquids are at constant volume and at 25°C, except as noted.

Note that [\[link\]](#) is an illustration of the mechanical equivalent of heat. Alternatively, the temperature increase could be produced by a blow torch instead of mechanically.

Example:

Calculating the Final Temperature When Heat Is Transferred Between Two Bodies: Pouring Cold Water in a Hot Pan

Suppose you pour 0.250 kg of 20.0°C water (about a cup) into a 0.500-kg aluminum pan off the stove with a temperature of 150°C. Assume that the pan is placed on an insulated pad and that a negligible amount of water boils off. What is the temperature when the water and pan reach thermal equilibrium a short time later?

Strategy

The pan is placed on an insulated pad so that little heat transfer occurs with the surroundings. Originally the pan and water are not in thermal equilibrium: the pan is at a higher temperature than the water. Heat transfer then restores thermal equilibrium once the water and pan are in contact. Because heat transfer between the pan and water takes place rapidly, the mass of evaporated water is negligible and the magnitude of the heat lost by the pan is equal to the heat gained by the water. The exchange of heat stops once a thermal equilibrium between the pan and the water is achieved. The heat exchange can be written as $|Q_{\text{hot}}| = Q_{\text{cold}}$.

Solution

1. Use the equation for heat transfer $Q = mc\Delta T$ to express the heat lost by the aluminum pan in terms of the mass of the pan, the specific heat of aluminum, the initial temperature of the pan, and the final temperature:

Equation:

$$Q_{\text{hot}} = m_{\text{Al}}c_{\text{Al}}(T_{\text{f}} - 150^{\circ}\text{C}).$$

2. Express the heat gained by the water in terms of the mass of the water, the specific heat of water, the initial temperature of the water and the final temperature:

Equation:

$$Q_{\text{cold}} = m_{\text{W}}c_{\text{W}}(T_{\text{f}} - 20.0^{\circ}\text{C}).$$

3. Note that $Q_{\text{hot}} < 0$ and $Q_{\text{cold}} > 0$ and that they must sum to zero because the heat lost by the hot pan must be the same as the heat gained by the cold water:

Equation:

$$\begin{aligned}Q_{\text{cold}} + Q_{\text{hot}} &= 0, \\Q_{\text{cold}} &= -Q_{\text{hot}}, \\m_{\text{W}}c_{\text{W}}(T_{\text{f}} - 20.0^{\circ}\text{C}) &= -m_{\text{Al}}c_{\text{Al}}(T_{\text{f}} - 150^{\circ}\text{C}).\end{aligned}$$

4. This an equation for the unknown final temperature, T_{f}
5. Bring all terms involving T_{f} on the left hand side and all other terms on the right hand side. Solve for T_{f} ,

Equation:

$$T_{\text{f}} = \frac{m_{\text{Al}}c_{\text{Al}}(150^{\circ}\text{C}) + m_{\text{W}}c_{\text{W}}(20.0^{\circ}\text{C})}{m_{\text{Al}}c_{\text{Al}} + m_{\text{W}}c_{\text{W}}},$$

and insert the numerical values:

Equation:

$$\begin{aligned}T_{\text{f}} &= \frac{(0.500 \text{ kg})(900 \text{ J/kg}^{\circ}\text{C})(150^{\circ}\text{C}) + (0.250 \text{ kg})(4186 \text{ J/kg}^{\circ}\text{C})(20.0^{\circ}\text{C})}{(0.500 \text{ kg})(900 \text{ J/kg}^{\circ}\text{C}) + (0.250 \text{ kg})(4186 \text{ J/kg}^{\circ}\text{C})} \\&= \frac{88430 \text{ J}}{1496.5 \text{ J}^{\circ}\text{C}} \\&= 59.1^{\circ}\text{C}.\end{aligned}$$

Discussion

This is a typical *calorimetry* problem—two bodies at different temperatures are brought in contact with each other and exchange heat until a common temperature is reached. Why is the final temperature so much closer to 20.0°C than 150°C ? The reason is that water has a greater specific heat than most common substances and thus undergoes a small temperature change for a given heat transfer. A large body of water, such as a lake, requires a large amount of heat to increase its temperature appreciably. This explains why the temperature of a lake stays relatively constant during a day even when the temperature change of the air is large. However, the water temperature does change over longer times (e.g., summer to winter).

Note:**Take-Home Experiment: Temperature Change of Land and Water**

What heats faster, land or water?

To study differences in heat capacity:

- Place equal masses of dry sand (or soil) and water at the same temperature into two small jars. (The average density of soil or sand is about 1.6 times that of water, so you can achieve approximately equal masses by using 50% more water by volume.)
- Heat both (using an oven or a heat lamp) for the same amount of time.
- Record the final temperature of the two masses.
- Now bring both jars to the same temperature by heating for a longer period of time.
- Remove the jars from the heat source and measure their temperature every 5 minutes for about 30 minutes.

Which sample cools off the fastest? This activity replicates the phenomena responsible for land breezes and sea breezes.

Exercise:**Check Your Understanding****Problem:**

If 25 kJ is necessary to raise the temperature of a block from 25°C to 30°C, how much heat is necessary to heat the block from 45°C to 50°C?

Solution:

The heat transfer depends only on the temperature difference. Since the temperature differences are the same in both cases, the same 25 kJ is necessary in the second case.

Summary

- The transfer of heat Q that leads to a change ΔT in the temperature of a body with mass m is $Q = mc\Delta T$, where c is the specific heat of the material. This relationship can also be considered as the definition of specific heat.

Conceptual Questions**Exercise:**

Problem:

What three factors affect the heat transfer that is necessary to change an object's temperature?

Exercise:**Problem:**

The brakes in a car increase in temperature by ΔT when bringing the car to rest from a speed v . How much greater would ΔT be if the car initially had twice the speed? You may assume the car to stop sufficiently fast so that no heat transfers out of the brakes.

Problems & Exercises**Exercise:****Problem:**

On a hot day, the temperature of an 80,000-L swimming pool increases by 1.50°C . What is the net heat transfer during this heating? Ignore any complications, such as loss of water by evaporation.

Solution:**Equation:**

$$5.02 \times 10^8 \text{ J}$$

Exercise:

Problem: Show that $1 \text{ cal/g} \cdot ^\circ\text{C} = 1 \text{ kcal/kg} \cdot ^\circ\text{C}$.

Exercise:**Problem:**

To sterilize a 50.0-g glass baby bottle, we must raise its temperature from 22.0°C to 95.0°C . How much heat transfer is required?

Solution:**Equation:**

$$3.07 \times 10^3 \text{ J}$$

Exercise:**Problem:**

The same heat transfer into identical masses of different substances produces different temperature changes. Calculate the final temperature when 1.00 kcal of heat transfers into 1.00 kg of the following, originally at 20.0°C: (a) water; (b) concrete; (c) steel; and (d) mercury.

Exercise:**Problem:**

Rubbing your hands together warms them by converting work into thermal energy. If a woman rubs her hands back and forth for a total of 20 rubs, at a distance of 7.50 cm per rub, and with an average frictional force of 40.0 N, what is the temperature increase? The mass of tissues warmed is only 0.100 kg, mostly in the palms and fingers.

Solution:**Equation:**

$$0.171^{\circ}\text{C}$$

Exercise:**Problem:**

A 0.250-kg block of a pure material is heated from 20.0°C to 65.0°C by the addition of 4.35 kJ of energy. Calculate its specific heat and identify the substance of which it is most likely composed.

Exercise:**Problem:**

Suppose identical amounts of heat transfer into different masses of copper and water, causing identical changes in temperature. What is the ratio of the mass of copper to water?

Solution:

$$10.8$$

Exercise:

Problem:

(a) The number of kilocalories in food is determined by calorimetry techniques in which the food is burned and the amount of heat transfer is measured. How many kilocalories per gram are there in a 5.00-g peanut if the energy from burning it is transferred to 0.500 kg of water held in a 0.100-kg aluminum cup, causing a 54.9°C temperature increase? (b) Compare your answer to labeling information found on a package of peanuts and comment on whether the values are consistent.

Exercise:**Problem:**

Following vigorous exercise, the body temperature of an 80.0-kg person is 40.0°C . At what rate in watts must the person transfer thermal energy to reduce the body temperature to 37.0°C in 30.0 min, assuming the body continues to produce energy at the rate of 150 W? (1 watt = 1 joule/second or $1\text{ W} = 1\text{ J/s}$).

Solution:

617 W

Exercise:**Problem:**

Even when shut down after a period of normal use, a large commercial nuclear reactor transfers thermal energy at the rate of 150 MW by the radioactive decay of fission products. This heat transfer causes a rapid increase in temperature if the cooling system fails

(1 watt = 1 joule/second or $1\text{ W} = 1\text{ J/s}$ and $1\text{ MW} = 1\text{ megawatt}$). (a) Calculate the rate of temperature increase in degrees Celsius per second ($^{\circ}\text{C/s}$) if the mass of the reactor core is $1.60 \times 10^5\text{ kg}$ and it has an average specific heat of $0.3349\text{ kJ/kg}^{\circ}\text{C}$. (b) How long would it take to obtain a temperature increase of 2000°C , which could cause some metals holding the radioactive materials to melt? (The initial rate of temperature increase would be greater than that calculated here because the heat transfer is concentrated in a smaller mass. Later, however, the temperature increase would slow down because the $5 \times 10^5\text{-kg}$ steel containment vessel would also begin to heat up.)



Radioactive spent-fuel pool at a nuclear power plant. Spent fuel stays hot for a long time. (credit: U.S. Department of Energy)

Glossary

specific heat

the amount of heat necessary to change the temperature of 1.00 kg of a substance by 1.00 °C

Phase Change and Latent Heat

- Examine heat transfer.
- Calculate final temperature from heat transfer.

So far we have discussed temperature change due to heat transfer. No temperature change occurs from heat transfer if ice melts and becomes liquid water (i.e., during a phase change). For example, consider water dripping from icicles melting on a roof warmed by the Sun. Conversely, water freezes in an ice tray cooled by lower-temperature surroundings.



Heat from the air transfers to the ice causing it to melt.
(credit: Mike Brand)

Energy is required to melt a solid because the cohesive bonds between the molecules in the solid must be broken apart such that, in the liquid, the molecules can move around at comparable kinetic energies; thus, there is no rise in temperature. Similarly, energy is needed to vaporize a liquid, because molecules in a liquid interact with each other via attractive forces. There is no temperature change until a phase change is complete. The temperature of a cup of soda initially at 0°C stays at 0°C until all the ice has melted. Conversely, energy is released during freezing and condensation, usually in the form of thermal energy. Work is done by cohesive forces when molecules are brought together. The corresponding energy must be given off (dissipated) to allow them to stay together [\[link\]](#).

The energy involved in a phase change depends on two major factors: the number and strength of bonds or force pairs. The number of bonds is proportional to the number of molecules and thus to the mass of the sample. The strength of forces depends on the type of molecules. The heat Q required to change the phase of a sample of mass m is given by

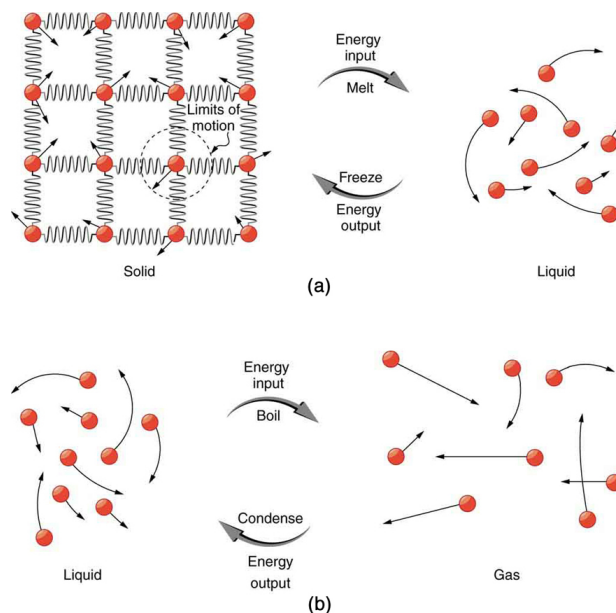
Equation:

$$Q = mL_f \text{ (melting/freezing),}$$

Equation:

$$Q = mL_v \text{ (vaporization/condensation),}$$

where the latent heat of fusion, L_f , and latent heat of vaporization, L_v , are material constants that are determined experimentally. See ([link](#)).



(a) Energy is required to partially overcome the attractive forces between molecules in a solid to form a liquid. That same energy must be removed for freezing to take place. (b) Molecules are separated by large distances when going from liquid to vapor, requiring significant energy to overcome molecular attraction. The same energy must be removed for condensation to take place. There is no temperature change until a phase change is complete.

Latent heat is measured in units of J/kg. Both L_f and L_v depend on the substance, particularly on the strength of its molecular forces as noted earlier. L_f and L_v are collectively called **latent heat coefficients**. They are *latent*, or hidden, because in phase changes, energy enters or leaves a system without causing a temperature change in the system; so, in effect, the energy is hidden. [\[link\]](#) lists representative values of L_f and L_v , together with melting and boiling points.

The table shows that significant amounts of energy are involved in phase changes. Let us look, for example, at how much energy is needed to melt a kilogram of ice at 0°C to produce a kilogram of water at 0°C . Using the equation for a change in temperature and the value for water from [\[link\]](#), we find that $Q = mL_f = (1.0\text{ kg})(334\text{ kJ/kg}) = 334\text{ kJ}$ is the energy to melt a kilogram of ice. This is a lot of energy as it represents the same amount of energy needed to raise the temperature of 1 kg of liquid water from 0°C to 79.8°C . Even more energy is required to vaporize water; it would take 2256 kJ to change 1 kg of liquid water at the normal boiling point (100°C at atmospheric pressure) to steam (water vapor). This example shows that the energy for a phase change is enormous compared to energy associated with temperature changes without a phase change.

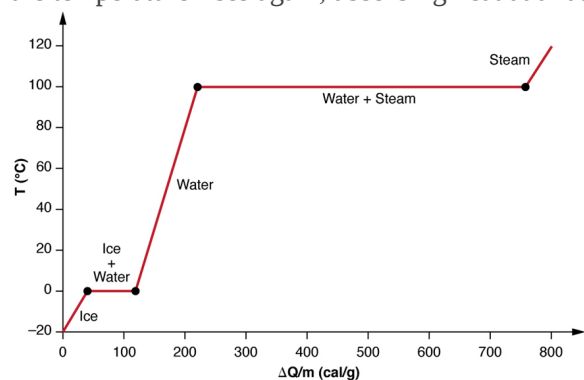
		L_f			L_v	
Substance	Melting point (°C)	kJ/kg	kcal/kg	Boiling point (°C)	kJ/kg	kcal/kg
Helium	−269.7	5.23	1.25	−268.9	20.9	4.99
Hydrogen	−259.3	58.6	14.0	−252.9	452	108
Nitrogen	−210.0	25.5	6.09	−195.8	201	48.0
Oxygen	−218.8	13.8	3.30	−183.0	213	50.9
Ethanol	−114	104	24.9	78.3	854	204
Ammonia	−75		108	−33.4	1370	327
Mercury	−38.9	11.8	2.82	357	272	65.0
Water	0.00	334	79.8	100.0	2256 ^[footnote] At 37.0°C (body temperature), the heat of vaporization L_v for water is 2430 kJ/kg or 580 kcal/kg	539 ^[footnote] At 37.0°C (body temperature), the heat of vaporization L_v for water is 2430 kJ/kg or 580 kcal/kg
Sulfur	119	38.1	9.10	444.6	326	77.9
Lead	327	24.5	5.85	1750	871	208
Antimony	631	165	39.4	1440	561	134
Aluminum	660	380	90	2450	11400	2720
Silver	961	88.3	21.1	2193	2336	558
Gold	1063	64.5	15.4	2660	1578	377
Copper	1083	134	32.0	2595	5069	1211
Uranium	1133	84	20	3900	1900	454
Tungsten	3410	184	44	5900	4810	1150

Heats of Fusion and Vaporization ^[footnote]

Values quoted at the normal melting and boiling temperatures at standard atmospheric pressure (1 atm).

Phase changes can have a tremendous stabilizing effect even on temperatures that are not near the melting and boiling points, because evaporation and condensation (conversion of a gas into a liquid state) occur even at temperatures below the boiling point. Take, for example, the fact that air temperatures in humid climates rarely go above 35.0°C , which is because most heat transfer goes into evaporating water into the air. Similarly, temperatures in humid weather rarely fall below the dew point because enormous heat is released when water vapor condenses.

We examine the effects of phase change more precisely by considering adding heat into a sample of ice at -20°C ([link](#)). The temperature of the ice rises linearly, absorbing heat at a constant rate of $0.50\text{ cal/g} \cdot ^{\circ}\text{C}$ until it reaches 0°C . Once at this temperature, the ice begins to melt until all the ice has melted, absorbing 79.8 cal/g of heat. The temperature remains constant at 0°C during this phase change. Once all the ice has melted, the temperature of the liquid water rises, absorbing heat at a new constant rate of $1.00\text{ cal/g} \cdot ^{\circ}\text{C}$. At 100°C , the water begins to boil and the temperature again remains constant while the water absorbs 539 cal/g of heat during this phase change. When all the liquid has become steam vapor, the temperature rises again, absorbing heat at a rate of $0.482\text{ cal/g} \cdot ^{\circ}\text{C}$.



A graph of temperature versus energy added. The system is constructed so that no vapor evaporates while ice warms to become liquid water, and so that, when vaporization occurs, the vapor remains in of the system. The long stretches of constant temperature values at 0°C and 100°C reflect the large latent heat of melting and vaporization, respectively.

Water can evaporate at temperatures below the boiling point. More energy is required than at the boiling point, because the kinetic energy of water molecules at temperatures below 100°C is less than that at 100°C , hence less energy is available from random thermal motions. Take, for example, the fact that, at body temperature, perspiration from the skin requires a heat input of 2428 kJ/kg , which is about 10 percent higher than the latent heat of vaporization at 100°C . This heat comes from the skin, and thus provides an effective cooling mechanism in hot weather. High humidity inhibits evaporation, so that body temperature might rise, leaving unevaporated sweat on your brow.

Example:
Calculate Final Temperature from Phase Change: Cooling Soda with Ice Cubes

Three ice cubes are used to chill a soda at 20°C with mass $m_{\text{soda}} = 0.25 \text{ kg}$. The ice is at 0°C and each ice cube has a mass of 6.0 g. Assume that the soda is kept in a foam container so that heat loss can be ignored. Assume the soda has the same heat capacity as water. Find the final temperature when all ice has melted.

Strategy

The ice cubes are at the melting temperature of 0°C. Heat is transferred from the soda to the ice for melting. Melting of ice occurs in two steps: first the phase change occurs and solid (ice) transforms into liquid water at the melting temperature, then the temperature of this water rises. Melting yields water at 0°C, so more heat is transferred from the soda to this water until the water plus soda system reaches thermal equilibrium,

Equation:

$$Q_{\text{ice}} = -Q_{\text{soda}}.$$

The heat transferred to the ice is $Q_{\text{ice}} = m_{\text{ice}}L_f + m_{\text{ice}}c_W(T_f - 0^\circ\text{C})$. The heat given off by the soda is $Q_{\text{soda}} = m_{\text{soda}}c_W(T_f - 20^\circ\text{C})$. Since no heat is lost, $Q_{\text{ice}} = -Q_{\text{soda}}$, so that

Equation:

$$m_{\text{ice}}L_f + m_{\text{ice}}c_W(T_f - 0^\circ\text{C}) = -m_{\text{soda}}c_W(T_f - 20^\circ\text{C}).$$

Bring all terms involving T_f on the left-hand-side and all other terms on the right-hand-side. Solve for the unknown quantity T_f :

Equation:

$$T_f = \frac{m_{\text{soda}}c_W(20^\circ\text{C}) - m_{\text{ice}}L_f}{(m_{\text{soda}} + m_{\text{ice}})c_W}.$$

Solution

1. Identify the known quantities. The mass of ice is $m_{\text{ice}} = 3 \times 6.0 \text{ g} = 0.018 \text{ kg}$ and the mass of soda is $m_{\text{soda}} = 0.25 \text{ kg}$.

2. Calculate the terms in the numerator:

Equation:

$$m_{\text{soda}}c_W(20^\circ\text{C}) = (0.25 \text{ kg})(4186 \text{ J/kg} \cdot ^\circ\text{C})(20^\circ\text{C}) = 20,930 \text{ J}$$

and

Equation:

$$m_{\text{ice}}L_f = (0.018 \text{ kg})(334,000 \text{ J/kg}) = 6012 \text{ J}.$$

3. Calculate the denominator:

Equation:

$$(m_{\text{soda}} + m_{\text{ice}})c_W = (0.25 \text{ kg} + 0.018 \text{ kg})(4186 \text{ J/(kg} \cdot ^\circ\text{C)}) = 1122 \text{ J/}^\circ\text{C}.$$

4. Calculate the final temperature:

Equation:

$$T_f = \frac{20,930 \text{ J} - 6012 \text{ J}}{1122 \text{ J/}^\circ\text{C}} = 13^\circ\text{C}.$$

Discussion

This example illustrates the enormous energies involved during a phase change. The mass of ice is about 7 percent the mass of water but leads to a noticeable change in the temperature of soda. Although we assumed that the ice was at the freezing temperature, this is incorrect: the typical temperature is -6°C . However, this correction gives a final temperature that is essentially identical to the result we found. Can you explain why?

We have seen that vaporization requires heat transfer to a liquid from the surroundings, so that energy is released by the surroundings. Condensation is the reverse process, increasing the temperature of the surroundings. This increase may seem surprising, since we associate condensation with cold objects—the glass in the figure, for example. However, energy must be removed from the condensing molecules to make a vapor condense. The energy is exactly the same as that required to make the phase change in the other direction, from liquid to vapor, and so it can be calculated from $Q = mL_v$.



Condensation forms on this glass of iced tea because the temperature of the nearby air is reduced to below the dew point. The rate at which water molecules join together exceeds the rate at which they separate, and so water condenses. Energy is released when the water condenses, speeding the melting of the ice in the glass. (credit: Jenny Downing)

Note:

Real-World Application

Energy is also released when a liquid freezes. This phenomenon is used by fruit growers in Florida to protect oranges when the temperature is close to the freezing point (0°C). Growers spray water on the

plants in orchards so that the water freezes and heat is released to the growing oranges on the trees. This prevents the temperature inside the orange from dropping below freezing, which would damage the fruit.



The ice on these trees released large amounts of energy when it froze, helping to prevent the temperature of the trees from dropping below 0°C .

Water is intentionally sprayed on orchards to help prevent hard frosts.

(credit: Hermann Hammer)

Sublimation is the transition from solid to vapor phase. You may have noticed that snow can disappear into thin air without a trace of liquid water, or the disappearance of ice cubes in a freezer. The reverse is also true: Frost can form on very cold windows without going through the liquid stage. A popular effect is the making of “smoke” from dry ice, which is solid carbon dioxide. Sublimation occurs because the equilibrium vapor pressure of solids is not zero. Certain air fresheners use the sublimation of a solid to inject a perfume into the room. Moth balls are a slightly toxic example of a phenol (an organic compound) that sublimates, while some solids, such as osmium tetroxide, are so toxic that they must be kept in sealed containers to prevent human exposure to their sublimation-produced vapors.



(a)



(b)

Direct transitions
between solid and

vapor are common,
sometimes useful,
and even beautiful.

(a) Dry ice
sublimates directly to
carbon dioxide gas.
The visible vapor is
made of water
droplets. (credit:
Windell Oskay) (b)
Frost forms patterns
on a very cold
window, an example
of a solid formed
directly from a
vapor. (credit: Liz
West)

All phase transitions involve heat. In the case of direct solid-vapor transitions, the energy required is given by the equation $Q = mL_s$, where L_s is the **heat of sublimation**, which is the energy required to change 1.00 kg of a substance from the solid phase to the vapor phase. L_s is analogous to L_f and L_v , and its value depends on the substance. Sublimation requires energy input, so that dry ice is an effective coolant, whereas the reverse process (i.e., frosting) releases energy. The amount of energy required for sublimation is of the same order of magnitude as that for other phase transitions.

The material presented in this section and the preceding section allows us to calculate any number of effects related to temperature and phase change. In each case, it is necessary to identify which temperature and phase changes are taking place and then to apply the appropriate equation. Keep in mind that heat transfer and work can cause both temperature and phase changes.

Problem-Solving Strategies for the Effects of Heat Transfer

1. *Examine the situation to determine that there is a change in the temperature or phase. Is there heat transfer into or out of the system?* When the presence or absence of a phase change is not obvious, you may wish to first solve the problem as if there were no phase changes, and examine the temperature change obtained. If it is sufficient to take you past a boiling or melting point, you should then go back and do the problem in steps—temperature change, phase change, subsequent temperature change, and so on.
2. *Identify and list all objects that change temperature and phase.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns).* A written list is useful.
4. *Make a list of what is given or what can be inferred from the problem as stated (identify the knowns).*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).* If there is a temperature change, the transferred heat depends on the specific heat (see [\[link\]](#)) whereas, for a phase change, the transferred heat depends on the latent heat. See [\[link\]](#).
6. *Substitute the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units.* You will need to do this in steps if there is more than one stage to the process (such as a temperature change followed by a phase change).

7. Check the answer to see if it is reasonable: Does it make sense? As an example, be certain that the temperature change does not also cause a phase change that you have not taken into account.

Exercise:

Check Your Understanding

Problem:

Why does snow remain on mountain slopes even when daytime temperatures are higher than the freezing temperature?

Solution:

Snow is formed from ice crystals and thus is the solid phase of water. Because enormous heat is necessary for phase changes, it takes a certain amount of time for this heat to be accumulated from the air, even if the air is above 0°C. The warmer the air is, the faster this heat exchange occurs and the faster the snow melts.

Summary

- Most substances can exist either in solid, liquid, and gas forms, which are referred to as “phases.”
- Phase changes occur at fixed temperatures for a given substance at a given pressure, and these temperatures are called boiling and freezing (or melting) points.
- During phase changes, heat absorbed or released is given by:

Equation:

$$Q = mL,$$

where L is the latent heat coefficient.

Conceptual Questions

Exercise:

Problem:

Heat transfer can cause temperature and phase changes. What else can cause these changes?

Exercise:

Problem:

How does the latent heat of fusion of water help slow the decrease of air temperatures, perhaps preventing temperatures from falling significantly below 0°C, in the vicinity of large bodies of water?

Exercise:

Problem: What is the temperature of ice right after it is formed by freezing water?

Exercise:

Problem:

If you place 0°C ice into 0°C water in an insulated container, what will happen? Will some ice melt, will more water freeze, or will neither take place?

Exercise:**Problem:**

What effect does condensation on a glass of ice water have on the rate at which the ice melts? Will the condensation speed up the melting process or slow it down?

Exercise:**Problem:**

In very humid climates where there are numerous bodies of water, such as in Florida, it is unusual for temperatures to rise above about 35°C (95°F). In deserts, however, temperatures can rise far above this. Explain how the evaporation of water helps limit high temperatures in humid climates.

Exercise:**Problem:**

In winters, it is often warmer in San Francisco than in nearby Sacramento, 150 km inland. In summers, it is nearly always hotter in Sacramento. Explain how the bodies of water surrounding San Francisco moderate its extreme temperatures.

Exercise:**Problem:**

Putting a lid on a boiling pot greatly reduces the heat transfer necessary to keep it boiling. Explain why.

Exercise:**Problem:**

Freeze-dried foods have been dehydrated in a vacuum. During the process, the food freezes and must be heated to facilitate dehydration. Explain both how the vacuum speeds up dehydration and why the food freezes as a result.

Exercise:**Problem:**

When still air cools by radiating at night, it is unusual for temperatures to fall below the dew point. Explain why.

Exercise:**Problem:**

In a physics classroom demonstration, an instructor inflates a balloon by mouth and then cools it in liquid nitrogen. When cold, the shrunken balloon has a small amount of light blue liquid in it, as well as some snow-like crystals. As it warms up, the liquid boils, and part of the crystals sublimate, with some crystals lingering for awhile and then producing a liquid. Identify the blue liquid and the two solids in the cold balloon. Justify your identifications using data from [\[link\]](#).

Problems & Exercises

Exercise:

Problem:

How much heat transfer (in kilocalories) is required to thaw a 0.450-kg package of frozen vegetables originally at 0°C if their heat of fusion is the same as that of water?

Solution:

35.9 kcal

Exercise:

Problem:

A bag containing 0°C ice is much more effective in absorbing energy than one containing the same amount of 0°C water.

- How much heat transfer is necessary to raise the temperature of 0.800 kg of water from 0°C to 30.0°C ?
- How much heat transfer is required to first melt 0.800 kg of 0°C ice and then raise its temperature?
- Explain how your answer supports the contention that the ice is more effective.

Exercise:

Problem:

- How much heat transfer is required to raise the temperature of a 0.750-kg aluminum pot containing 2.50 kg of water from 30.0°C to the boiling point and then boil away 0.750 kg of water?
 - How long does this take if the rate of heat transfer is 500 W
 $1 \text{ watt} = 1 \text{ joule/second}$ ($1 \text{ W} = 1 \text{ J/s}$)?
-

Solution:

(a) 591 kcal

(b) $4.94 \times 10^3 \text{ s}$

Exercise:

Problem:

The formation of condensation on a glass of ice water causes the ice to melt faster than it would otherwise. If 8.00 g of condensation forms on a glass containing both water and 200 g of ice, how many grams of the ice will melt as a result? Assume no other heat transfer occurs.

Exercise:

Problem:

On a trip, you notice that a 3.50-kg bag of ice lasts an average of one day in your cooler. What is the average power in watts entering the ice if it starts at 0°C and completely melts to 0°C water in exactly one day $1 \text{ watt} = 1 \text{ joule/second}$ ($1 \text{ W} = 1 \text{ J/s}$)?

Solution:

13.5 W

Exercise:**Problem:**

On a certain dry sunny day, a swimming pool's temperature would rise by 1.50°C if not for evaporation. What fraction of the water must evaporate to carry away precisely enough energy to keep the temperature constant?

Exercise:**Problem:**

- (a) How much heat transfer is necessary to raise the temperature of a 0.200-kg piece of ice from -20.0°C to 130°C , including the energy needed for phase changes?
- (b) How much time is required for each stage, assuming a constant 20.0 kJ/s rate of heat transfer?
- (c) Make a graph of temperature versus time for this process.

Solution:

(a) 148 kcal

(b) 0.418 s, 3.34 s, 4.19 s, 22.6 s, 0.456 s

Exercise:**Problem:**

In 1986, a gargantuan iceberg broke away from the Ross Ice Shelf in Antarctica. It was approximately a rectangle 160 km long, 40.0 km wide, and 250 m thick.

- (a) What is the mass of this iceberg, given that the density of ice is 917 kg/m^3 ?
- (b) How much heat transfer (in joules) is needed to melt it?
- (c) How many years would it take sunlight alone to melt ice this thick, if the ice absorbs an average of 100 W/m^2 , 12.00 h per day?

Exercise:**Problem:**

How many grams of coffee must evaporate from 350 g of coffee in a 100-g glass cup to cool the coffee from 95.0°C to 45.0°C ? You may assume the coffee has the same thermal properties as water and that the average heat of vaporization is 2340 kJ/kg (560 cal/g). (You may neglect the change in mass of the coffee as it cools, which will give you an answer that is slightly larger than correct.)

Solution:

33.0 g

Exercise:

Problem:

- (a) It is difficult to extinguish a fire on a crude oil tanker, because each liter of crude oil releases 2.80×10^7 J of energy when burned. To illustrate this difficulty, calculate the number of liters of water that must be expended to absorb the energy released by burning 1.00 L of crude oil, if the water has its temperature raised from 20.0°C to 100°C , it boils, and the resulting steam is raised to 300°C .
- (b) Discuss additional complications caused by the fact that crude oil has a smaller density than water.
-

Solution:

(a) 9.67 L

(b) Crude oil is less dense than water, so it floats on top of the water, thereby exposing it to the oxygen in the air, which it uses to burn. Also, if the water is under the oil, it is less efficient in absorbing the heat generated by the oil.

Exercise:**Problem:**

The energy released from condensation in thunderstorms can be very large. Calculate the energy released into the atmosphere for a small storm of radius 1 km, assuming that 1.0 cm of rain is precipitated uniformly over this area.

Exercise:

Problem: To help prevent frost damage, 4.00 kg of 0°C water is sprayed onto a fruit tree.

- (a) How much heat transfer occurs as the water freezes?
- (b) How much would the temperature of the 200-kg tree decrease if this amount of heat transferred from the tree? Take the specific heat to be $3.35 \text{ kJ/kg} \cdot ^\circ\text{C}$, and assume that no phase change occurs.
-

Solution:

a) 319 kcal

b) 2.00°C

Exercise:**Problem:**

A 0.250-kg aluminum bowl holding 0.800 kg of soup at 25.0°C is placed in a freezer. What is the final temperature if 377 kJ of energy is transferred from the bowl and soup, assuming the soup's thermal properties are the same as that of water? Explicitly show how you follow the steps in [Problem-Solving Strategies for the Effects of Heat Transfer](#).

Exercise:**Problem:**

A 0.0500-kg ice cube at -30.0°C is placed in 0.400 kg of 35.0°C water in a very well-insulated container. What is the final temperature?

Solution:

20.6°C

Exercise:**Problem:**

If you pour 0.0100 kg of 20.0°C water onto a 1.20-kg block of ice (which is initially at -15.0°C), what is the final temperature? You may assume that the water cools so rapidly that effects of the surroundings are negligible.

Exercise:**Problem:**

Indigenous people sometimes cook in watertight baskets by placing hot rocks into water to bring it to a boil. What mass of 500°C rock must be placed in 4.00 kg of 15.0°C water to bring its temperature to 100°C, if 0.0250 kg of water escapes as vapor from the initial sizzle? You may neglect the effects of the surroundings and take the average specific heat of the rocks to be that of granite.

Solution:

4.38 kg

Exercise:**Problem:**

What would be the final temperature of the pan and water in [Calculating the Final Temperature When Heat Is Transferred Between Two Bodies: Pouring Cold Water in a Hot Pan](#) if 0.260 kg of water was placed in the pan and 0.0100 kg of the water evaporated immediately, leaving the remainder to come to a common temperature with the pan?

Exercise:**Problem:**

In some countries, liquid nitrogen is used on dairy trucks instead of mechanical refrigerators. A 3.00-hour delivery trip requires 200 L of liquid nitrogen, which has a density of 808 kg/m^3 .

(a) Calculate the heat transfer necessary to evaporate this amount of liquid nitrogen and raise its temperature to 3.00°C. (Use c_p and assume it is constant over the temperature range.) This value is the amount of cooling the liquid nitrogen supplies.

(b) What is this heat transfer rate in kilowatt-hours?

(c) Compare the amount of cooling obtained from melting an identical mass of 0°C ice with that from evaporating the liquid nitrogen.

Solution:

(a) $1.57 \times 10^4\text{ kcal}$

(b) 18.3 kW · h

(c) $1.29 \times 10^4\text{ kcal}$

Exercise:**Problem:**

Some gun fanciers make their own bullets, which involves melting and casting the lead slugs. How much heat transfer is needed to raise the temperature and melt 0.500 kg of lead, starting from 25.0°C ?

Glossary

heat of sublimation

the energy required to change a substance from the solid phase to the vapor phase

latent heat coefficient

a physical constant equal to the amount of heat transferred for every 1 kg of a substance during the change in phase of the substance

sublimation

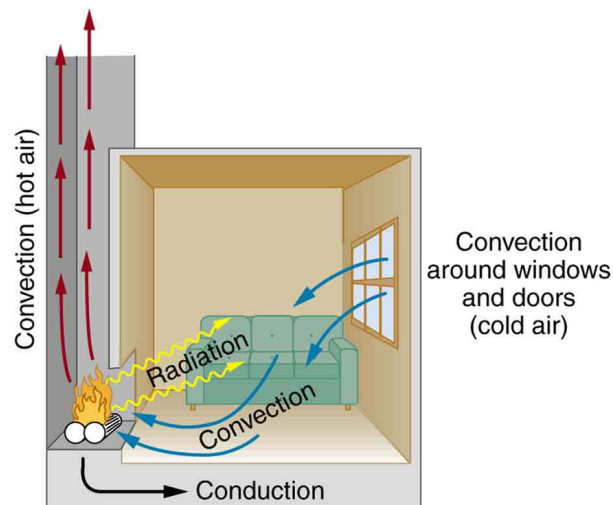
the transition from the solid phase to the vapor phase

Heat Transfer Methods

- Discuss the different methods of heat transfer.

Equally as interesting as the effects of heat transfer on a system are the methods by which this occurs. Whenever there is a temperature difference, heat transfer occurs. Heat transfer may occur rapidly, such as through a cooking pan, or slowly, such as through the walls of a picnic ice chest. We can control rates of heat transfer by choosing materials (such as thick wool clothing for the winter), controlling air movement (such as the use of weather stripping around doors), or by choice of color (such as a white roof to reflect summer sunlight). So many processes involve heat transfer, so that it is hard to imagine a situation where no heat transfer occurs. Yet every process involving heat transfer takes place by only three methods:

1. **Conduction** is heat transfer through stationary matter by physical contact. (The matter is stationary on a macroscopic scale—we know there is thermal motion of the atoms and molecules at any temperature above absolute zero.) Heat transferred between the electric burner of a stove and the bottom of a pan is transferred by conduction.
2. **Convection** is the heat transfer by the macroscopic movement of a fluid. This type of transfer takes place in a forced-air furnace and in weather systems, for example.
3. Heat transfer by **radiation** occurs when microwaves, infrared radiation, visible light, or another form of electromagnetic radiation is emitted or absorbed. An obvious example is the warming of the Earth by the Sun. A less obvious example is thermal radiation from the human body.



In a fireplace, heat transfer occurs by all three methods: conduction, convection, and radiation. Radiation is responsible for most of the heat transferred into the room. Heat transfer also occurs through conduction into the room, but at a much slower rate. Heat transfer by convection also occurs through cold air entering the room around windows and hot air leaving the room by rising up the chimney.

We examine these methods in some detail in the three following modules. Each method has unique and interesting characteristics, but all three do have one thing in common: they transfer heat solely because of a temperature difference [\[link\]](#).

Exercise:

Check Your Understanding

Problem:

Name an example from daily life (different from the text) for each mechanism of heat transfer.

Solution:

Conduction: Heat transfers into your hands as you hold a hot cup of coffee.

Convection: Heat transfers as the barista “steams” cold milk to make hot *cocoa*.

Radiation: Reheating a cold cup of coffee in a microwave oven.

Summary

- Heat is transferred by three different methods: conduction, convection, and radiation.

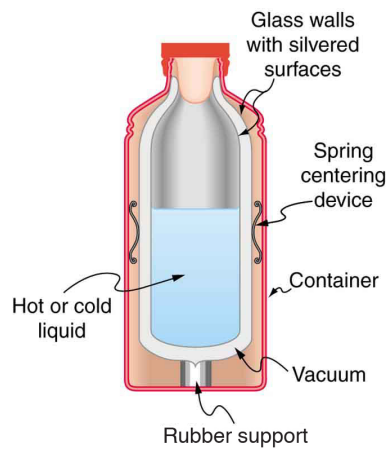
Conceptual Questions**Exercise:****Problem:**

What are the main methods of heat transfer from the hot core of Earth to its surface? From Earth’s surface to outer space?

When our bodies get too warm, they respond by sweating and increasing blood circulation to the surface to transfer thermal energy away from the core. What effect will this have on a person in a ° hot tub?

[\[link\]](#) shows a cut-away drawing of a thermos bottle (also known as a Dewar flask), which is a device designed specifically to slow down all forms of heat transfer. Explain the functions of the various parts, such as the

vacuum, the silvering of the walls, the thin-walled long glass neck, the rubber support, the air layer, and the stopper.



The construction of a thermos bottle is designed to inhibit all methods of heat transfer.

Glossary

conduction

heat transfer through stationary matter by physical contact

convection

heat transfer by the macroscopic movement of fluid

radiation

heat transfer which occurs when microwaves, infrared radiation, visible light, or other electromagnetic radiation is emitted or absorbed

Conduction

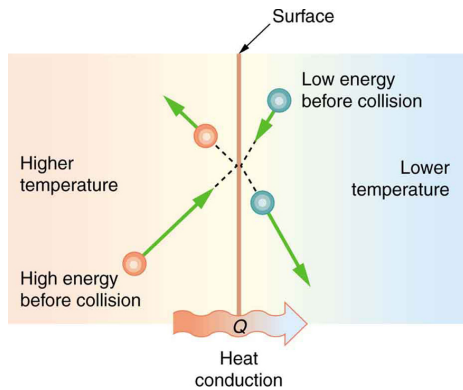
- Calculate thermal conductivity.
- Observe conduction of heat in collisions.
- Study thermal conductivities of common substances.



Insulation is used to limit the conduction of heat from the inside to the outside (in winters) and from the outside to the inside (in summers).
(credit: Giles Douglas)

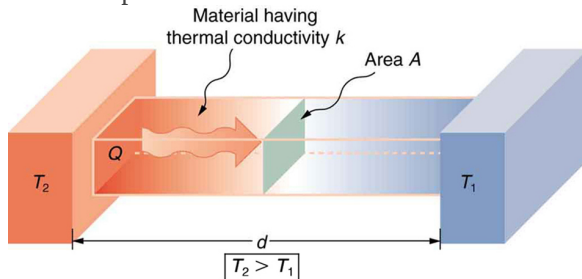
Your feet feel cold as you walk barefoot across the living room carpet in your cold house and then step onto the kitchen tile floor. This result is intriguing, since the carpet and tile floor are both at the same temperature. The different sensation you feel is explained by the different rates of heat transfer: the heat loss during the same time interval is greater for skin in contact with the tiles than with the carpet, so the temperature drop is greater on the tiles.

Some materials conduct thermal energy faster than others. In general, good conductors of electricity (metals like copper, aluminum, gold, and silver) are also good heat conductors, whereas insulators of electricity (wood, plastic, and rubber) are poor heat conductors. [\[link\]](#) shows molecules in two bodies at different temperatures. The (average) kinetic energy of a molecule in the hot body is higher than in the colder body. If two molecules collide, an energy transfer from the molecule with greater kinetic energy to the molecule with less kinetic energy occurs. The cumulative effect from all collisions results in a net flux of heat from the hot body to the colder body. The heat flux thus depends on the temperature difference $\Delta T = T_{\text{hot}} - T_{\text{cold}}$. Therefore, you will get a more severe burn from boiling water than from hot tap water. Conversely, if the temperatures are the same, the net heat transfer rate falls to zero, and equilibrium is achieved. Owing to the fact that the number of collisions increases with increasing area, heat conduction depends on the cross-sectional area. If you touch a cold wall with your palm, your hand cools faster than if you just touch it with your fingertip.



The molecules in two bodies at different temperatures have different average kinetic energies. Collisions occurring at the contact surface tend to transfer energy from high-temperature regions to low-temperature regions. In this illustration, a molecule in the lower temperature region (right side) has low energy before collision, but its energy increases after colliding with the contact surface. In contrast, a molecule in the higher temperature region (left side) has high energy before collision, but its energy decreases after colliding with the contact surface.

A third factor in the mechanism of conduction is the thickness of the material through which heat transfers. The figure below shows a slab of material with different temperatures on either side. Suppose that T_2 is greater than T_1 , so that heat is transferred from left to right. Heat transfer from the left side to the right side is accomplished by a series of molecular collisions. The thicker the material, the more time it takes to transfer the same amount of heat. This model explains why thick clothing is warmer than thin clothing in winters, and why Arctic mammals protect themselves with thick blubber.



Heat conduction occurs through any material, represented here by a rectangular bar, whether window glass or walrus blubber. The temperature of the material is T_2 on the left and T_1 on the right, where T_2 is greater than T_1 .

The rate of heat transfer by conduction is directly proportional to the surface area A , the temperature difference $T_2 - T_1$, and the substance's conductivity k . The rate of heat transfer is inversely proportional to the thickness d .

Lastly, the heat transfer rate depends on the material properties described by the coefficient of thermal conductivity. All four factors are included in a simple equation that was deduced from and is confirmed by experiments. The **rate of conductive heat transfer** through a slab of material, such as the one in [\[link\]](#), is given by

Equation:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d},$$

where Q/t is the rate of heat transfer in watts or kilocalories per second, k is the **thermal conductivity** of the material, A and d are its surface area and thickness, as shown in [\[link\]](#), and $(T_2 - T_1)$ is the temperature difference across the slab. [\[link\]](#) gives representative values of thermal conductivity.

Example:

Calculating Heat Transfer Through Conduction: Conduction Rate Through an Ice Box

A Styrofoam ice box has a total area of 0.950 m^2 and walls with an average thickness of 2.50 cm . The box contains ice, water, and canned beverages at 0°C . The inside of the box is kept cold by melting ice. How much ice melts in one day if the ice box is kept in the trunk of a car at 35.0°C ?

Strategy

This question involves both heat for a phase change (melting of ice) and the transfer of heat by conduction. To find the amount of ice melted, we must find the net heat transferred. This value can be obtained by calculating the rate of heat transfer by conduction and multiplying by time.

Solution

1. Identify the knowns.

Equation:

$$A = 0.950 \text{ m}^2; d = 2.50 \text{ cm} = 0.0250 \text{ m}; T_1 = 0^\circ\text{C}; T_2 = 35.0^\circ\text{C}; t = 1 \text{ day} = 24 \text{ hours} = 86,400 \text{ s}.$$

2. Identify the unknowns. We need to solve for the mass of the ice, m . We will also need to solve for the net heat transferred to melt the ice, Q .
3. Determine which equations to use. The rate of heat transfer by conduction is given by

Equation:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d}.$$

4. The heat is used to melt the ice: $Q = mL_f$.
5. Insert the known values:

Equation:

$$\frac{Q}{t} = \frac{(0.010 \text{ J/s} \cdot \text{m} \cdot ^\circ\text{C})(0.950 \text{ m}^2)(35.0^\circ\text{C} - 0^\circ\text{C})}{0.0250 \text{ m}} = 13.3 \text{ J/s}.$$

6. Multiply the rate of heat transfer by the time (1 day = 86,400 s):

Equation:

$$Q = (Q/t)t = (13.3 \text{ J/s})(86,400 \text{ s}) = 1.15 \times 10^6 \text{ J}.$$

7. Set this equal to the heat transferred to melt the ice: $Q = mL_f$. Solve for the mass m :

Equation:

$$m = \frac{Q}{L_f} = \frac{1.15 \times 10^6 \text{ J}}{334 \times 10^3 \text{ J/kg}} = 3.44 \text{ kg}.$$

Discussion

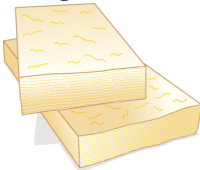
The result of 3.44 kg, or about 7.6 lbs, seems about right, based on experience. You might expect to use about a 4 kg (7–10 lb) bag of ice per day. A little extra ice is required if you add any warm food or beverages. Inspecting the conductivities in [\[link\]](#) shows that Styrofoam is a very poor conductor and thus a good insulator. Other good insulators include fiberglass, wool, and goose-down feathers. Like Styrofoam, these all incorporate many small pockets of air, taking advantage of air's poor thermal conductivity.

Substance	Thermal conductivity k (J/s·m·°C)
Silver	420
Copper	390
Gold	318
Aluminum	220
Steel iron	80
Steel (stainless)	14
Ice	2.2
Glass (average)	0.84
Concrete brick	0.84
Water	0.6
Fatty tissue (without blood)	0.2
Asbestos	0.16
Plasterboard	0.16
Wood	0.08–0.16

Substance	Thermal conductivity k (J/s·m·°C)
Snow (dry)	0.10
Cork	0.042
Glass wool	0.042
Wool	0.04
Down feathers	0.025
Air	0.023
Styrofoam	0.010

Thermal Conductivities of Common Substances[\[footnote\]](#)
At temperatures near 0°C.

A combination of material and thickness is often manipulated to develop good insulators—the smaller the conductivity k and the larger the thickness d , the better. The ratio of d/k will thus be large for a good insulator. The ratio d/k is called the **R factor**. The rate of conductive heat transfer is inversely proportional to R . The larger the value of R , the better the insulation. R factors are most commonly quoted for household insulation, refrigerators, and the like—unfortunately, it is still in non-metric units of $\text{ft}^2 \cdot ^\circ\text{F} \cdot \text{h} / \text{Btu}$, although the unit usually goes unstated (1 British thermal unit [Btu] is the amount of energy needed to change the temperature of 1.0 lb of water by 1.0 °F). A couple of representative values are an R factor of 11 for 3.5-in-thick fiberglass batts (pieces) of insulation and an R factor of 19 for 6.5-in-thick fiberglass batts. Walls are usually insulated with 3.5-in batts, while ceilings are usually insulated with 6.5-in batts. In cold climates, thicker batts may be used in ceilings and walls.



The fiberglass batt is used for insulation of walls and ceilings to prevent heat transfer between the inside of the building and the outside environment.

Note that in [\[link\]](#), the best thermal conductors—silver, copper, gold, and aluminum—are also the best electrical conductors, again related to the density of free electrons in them. Cooking utensils are typically made

from good conductors.

Example:

Calculating the Temperature Difference Maintained by a Heat Transfer: Conduction Through an Aluminum Pan

Water is boiling in an aluminum pan placed on an electrical element on a stovetop. The sauce pan has a bottom that is 0.800 cm thick and 14.0 cm in diameter. The boiling water is evaporating at the rate of 1.00 g/s. What is the temperature difference across (through) the bottom of the pan?

Strategy

Conduction through the aluminum is the primary method of heat transfer here, and so we use the equation for the rate of heat transfer and solve for the temperature difference.

Equation:

$$T_2 - T_1 = \frac{Q}{t} \left(\frac{d}{kA} \right).$$

Solution

1. Identify the knowns and convert them to the SI units.

The thickness of the pan, $d = 0.800 \text{ cm} = 8.0 \times 10^{-3} \text{ m}$, the area of the pan, $A = \pi(0.14/2)^2 \text{ m}^2 = 1.54 \times 10^{-2} \text{ m}^2$, and the thermal conductivity, $k = 220 \text{ J/s} \cdot \text{m} \cdot ^\circ\text{C}$.

2. Calculate the necessary heat of vaporization of 1 g of water:

Equation:

$$Q = mL_v = (1.00 \times 10^{-3} \text{ kg})(2256 \times 10^3 \text{ J/kg}) = 2256 \text{ J}.$$

3. Calculate the rate of heat transfer given that 1 g of water melts in one second:

Equation:

$$Q/t = 2256 \text{ J/s or } 2.26 \text{ kW}.$$

4. Insert the knowns into the equation and solve for the temperature difference:

Equation:

$$T_2 - T_1 = \frac{Q}{t} \left(\frac{d}{kA} \right) = (2256 \text{ J/s}) \frac{8.00 \times 10^{-3} \text{ m}}{(220 \text{ J/s} \cdot \text{m} \cdot ^\circ\text{C})(1.54 \times 10^{-2} \text{ m}^2)} = 5.33^\circ\text{C}.$$

Discussion

The value for the heat transfer $Q/t = 2.26 \text{ kW}$ or 2256 J/s is typical for an electric stove. This value gives a remarkably small temperature difference between the stove and the pan. Consider that the stove burner is red hot while the inside of the pan is nearly 100°C because of its contact with boiling water. This contact effectively cools the bottom of the pan in spite of its proximity to the very hot stove burner. Aluminum is such a good conductor that it only takes this small temperature difference to produce a heat transfer of 2.26 kW into the pan.

Conduction is caused by the random motion of atoms and molecules. As such, it is an ineffective mechanism for heat transport over macroscopic distances and short time distances. Take, for example, the temperature on the Earth, which would be unbearably cold during the night and extremely hot during the day if heat transport in the atmosphere was to be only through conduction. In another example, car engines would overheat unless there was a more efficient way to remove excess heat from the pistons.

Exercise:
Check Your Understanding

Problem:

How does the rate of heat transfer by conduction change when all spatial dimensions are doubled?

Solution:

Because area is the product of two spatial dimensions, it increases by a factor of four when each dimension is doubled ($A_{\text{final}} = (2d)^2 = 4d^2 = 4A_{\text{initial}}$). The distance, however, simply doubles. Because the temperature difference and the coefficient of thermal conductivity are independent of the spatial dimensions, the rate of heat transfer by conduction increases by a factor of four divided by two, or two:

Equation:

$$\left(\frac{Q}{t}\right)_{\text{final}} = \frac{kA_{\text{final}}(T_2 - T_1)}{d_{\text{final}}} = \frac{k(4A_{\text{initial}})(T_2 - T_1)}{2d_{\text{initial}}} = 2 \frac{kA_{\text{initial}}(T_2 - T_1)}{d_{\text{initial}}} = 2 \left(\frac{Q}{t}\right)_{\text{initial}}.$$

Summary

- Heat conduction is the transfer of heat between two objects in direct contact with each other.
- The rate of heat transfer Q/t (energy per unit time) is proportional to the temperature difference $T_2 - T_1$ and the contact area A and inversely proportional to the distance d between the objects:

Equation:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d}.$$

Conceptual Questions

Exercise:

Problem:

Some electric stoves have a flat ceramic surface with heating elements hidden beneath. A pot placed over a heating element will be heated, while it is safe to touch the surface only a few centimeters away. Why is ceramic, with a conductivity less than that of a metal but greater than that of a good insulator, an ideal choice for the stove top?

Exercise:

Problem:

Loose-fitting white clothing covering most of the body is ideal for desert dwellers, both in the hot Sun and during cold evenings. Explain how such clothing is advantageous during both day and night.



A jellabiya is worn by many men in Egypt. (credit: Zerida)

Problems & Exercises

Exercise:

Problem:

- (a) Calculate the rate of heat conduction through house walls that are 13.0 cm thick and that have an average thermal conductivity twice that of glass wool. Assume there are no windows or doors. The surface area of the walls is 120 m^2 and their inside surface is at 18.0°C , while their outside surface is at 5.00°C .
- (b) How many 1-kW room heaters would be needed to balance the heat transfer due to conduction?

Solution:

- (a) $1.01 \times 10^3 \text{ W}$
- (b) One

Exercise:

Problem:

The rate of heat conduction out of a window on a winter day is rapid enough to chill the air next to it. To see just how rapidly the windows transfer heat by conduction, calculate the rate of conduction in watts through a 3.00-m^2 window that is 0.635 cm thick (1/4 in) if the temperatures of the inner and outer surfaces are 5.00°C and -10.0°C , respectively. This rapid rate will not be maintained—the inner surface will cool, and even result in frost formation.

Exercise:

Problem:

Calculate the rate of heat conduction out of the human body, assuming that the core internal temperature is 37.0°C , the skin temperature is 34.0°C , the thickness of the tissues between averages 1.00 cm, and the surface area is 1.40 m^2 .

Solution:

84.0 W

Exercise:**Problem:**

Suppose you stand with one foot on ceramic flooring and one foot on a wool carpet, making contact over an area of 80.0 cm^2 with each foot. Both the ceramic and the carpet are 2.00 cm thick and are 10.0°C on their bottom sides. At what rate must heat transfer occur from each foot to keep the top of the ceramic and carpet at 33.0°C ?

Exercise:**Problem:**

A man consumes 3000 kcal of food in one day, converting most of it to maintain body temperature. If he loses half this energy by evaporating water (through breathing and sweating), how many kilograms of water evaporate?

Solution:

2.59 kg

Exercise:**Problem:**

(a) A firewalker runs across a bed of hot coals without sustaining burns. Calculate the heat transferred by conduction into the sole of one foot of a firewalker given that the bottom of the foot is a 3.00-mm-thick callus with a conductivity at the low end of the range for wood and its density is 300 kg/m^3 . The area of contact is 25.0 cm^2 , the temperature of the coals is 700°C , and the time in contact is 1.00 s.

(b) What temperature increase is produced in the 25.0 cm^3 of tissue affected?

(c) What effect do you think this will have on the tissue, keeping in mind that a callus is made of dead cells?

Exercise:**Problem:**

(a) What is the rate of heat conduction through the 3.00-cm-thick fur of a large animal having a 1.40-m^2 surface area? Assume that the animal's skin temperature is 32.0°C , that the air temperature is -5.00°C , and that fur has the same thermal conductivity as air. (b) What food intake will the animal need in one day to replace this heat transfer?

Solution:

(a) 39.7 W

(b) 820 kcal

Exercise:

Problem:

A walrus transfers energy by conduction through its blubber at the rate of 150 W when immersed in -1.00°C water. The walrus's internal core temperature is 37.0°C , and it has a surface area of 2.00 m^2 . What is the average thickness of its blubber, which has the conductivity of fatty tissues without blood?



Walrus on ice. (credit: Captain Budd Christman, NOAA Corps)

Exercise:**Problem:**

Compare the rate of heat conduction through a 13.0-cm-thick wall that has an area of 10.0 m^2 and a thermal conductivity twice that of glass wool with the rate of heat conduction through a window that is 0.750 cm thick and that has an area of 2.00 m^2 , assuming the same temperature difference across each.

Solution:

35 to 1, window to wall

Exercise:**Problem:**

Suppose a person is covered head to foot by wool clothing with average thickness of 2.00 cm and is transferring energy by conduction through the clothing at the rate of 50.0 W. What is the temperature difference across the clothing, given the surface area is 1.40 m^2 ?

Exercise:**Problem:**

Some stove tops are smooth ceramic for easy cleaning. If the ceramic is 0.600 cm thick and heat conduction occurs through the same area and at the same rate as computed in [\[link\]](#), what is the temperature difference across it? Ceramic has the same thermal conductivity as glass and brick.

Solution:

$1.05 \times 10^3\text{ K}$

Exercise:

Problem:

One easy way to reduce heating (and cooling) costs is to add extra insulation in the attic of a house. Suppose the house already had 15 cm of fiberglass insulation in the attic and in all the exterior surfaces. If you added an extra 8.0 cm of fiberglass to the attic, then by what percentage would the heating cost of the house drop? Take the single story house to be of dimensions 10 m by 15 m by 3.0 m. Ignore air infiltration and heat loss through windows and doors.

Exercise:**Problem:**

(a) Calculate the rate of heat conduction through a double-paned window that has a 1.50-m^2 area and is made of two panes of 0.800-cm-thick glass separated by a 1.00-cm air gap. The inside surface temperature is 15.0°C , while that on the outside is -10.0°C . (Hint: There are identical temperature drops across the two glass panes. First find these and then the temperature drop across the air gap. This problem ignores the increased heat transfer in the air gap due to convection.)

(b) Calculate the rate of heat conduction through a 1.60-cm-thick window of the same area and with the same temperatures. Compare your answer with that for part (a).

Solution:

(a) 83 W

(b) 24 times that of a double pane window.

Exercise:**Problem:**

Many decisions are made on the basis of the payback period: the time it will take through savings to equal the capital cost of an investment. Acceptable payback times depend upon the business or philosophy one has. (For some industries, a payback period is as small as two years.) Suppose you wish to install the extra insulation in [\[link\]](#). If energy cost \$1.00 per million joules and the insulation was \$4.00 per square meter, then calculate the simple payback time. Take the average ΔT for the 120 day heating season to be 15.0°C .

Exercise:**Problem:**

For the human body, what is the rate of heat transfer by conduction through the body's tissue with the following conditions: the tissue thickness is 3.00 cm, the change in temperature is 2.00°C , and the skin area is 1.50 m^2 . How does this compare with the average heat transfer rate to the body resulting from an energy intake of about 2400 kcal per day? (No exercise is included.)

Solution:

20.0 W, 17.2% of 2400 kcal per day

Glossary

R factor

the ratio of thickness to the conductivity of a material

rate of conductive heat transfer

rate of heat transfer from one material to another

thermal conductivity

the property of a material's ability to conduct heat

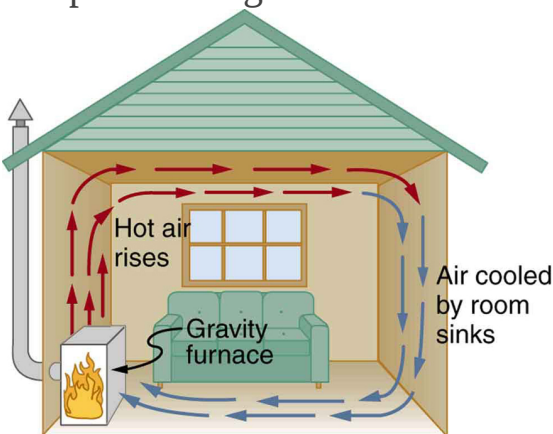
Convection

- Discuss the method of heat transfer by convection.

Convection is driven by large-scale flow of matter. In the case of Earth, the atmospheric circulation is caused by the flow of hot air from the tropics to the poles, and the flow of cold air from the poles toward the tropics. (Note that Earth's rotation causes the observed easterly flow of air in the northern hemisphere). Car engines are kept cool by the flow of water in the cooling system, with the water pump maintaining a flow of cool water to the pistons. The circulatory system is used the body: when the body overheats, the blood vessels in the skin expand (dilate), which increases the blood flow to the skin where it can be cooled by sweating. These vessels become smaller when it is cold outside and larger when it is hot (so more fluid flows, and more energy is transferred).

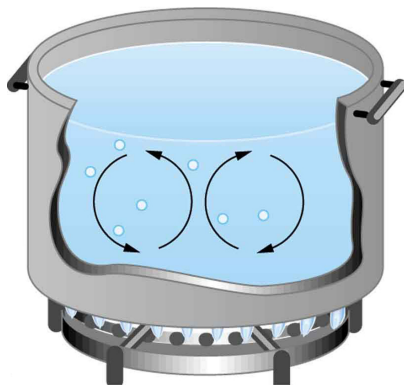
The body also loses a significant fraction of its heat through the breathing process.

While convection is usually more complicated than conduction, we can describe convection and do some straightforward, realistic calculations of its effects. Natural convection is driven by buoyant forces: hot air rises because density decreases as temperature increases. The house in [\[link\]](#) is kept warm in this manner, as is the pot of water on the stove in [\[link\]](#). Ocean currents and large-scale atmospheric circulation transfer energy from one part of the globe to another. Both are examples of natural convection.



Air heated by the so-called

gravity furnace expands and rises, forming a convective loop that transfers energy to other parts of the room. As the air is cooled at the ceiling and outside walls, it contracts, eventually becoming denser than room air and sinking to the floor. A properly designed heating system using natural convection, like this one, can be quite efficient in uniformly heating a home.



Convection plays an important role in heat transfer inside this pot of water. Once conducted to the inside, heat transfer to other parts of the pot is mostly by convection. The hotter water expands, decreases

in density, and rises to transfer heat to other regions of the water, while colder water sinks to the bottom. This process keeps repeating.

Note:

Take-Home Experiment: Convection Rolls in a Heated Pan

Take two small pots of water and use an eye dropper to place a drop of food coloring near the bottom of each. Leave one on a bench top and heat the other over a stovetop. Watch how the color spreads and how long it takes the color to reach the top. Watch how convective loops form.

Example:

Calculating Heat Transfer by Convection: Convection of Air Through the Walls of a House

Most houses are not airtight: air goes in and out around doors and windows, through cracks and crevices, following wiring to switches and outlets, and so on. The air in a typical house is completely replaced in less than an hour. Suppose that a moderately-sized house has inside dimensions $12.0\text{m} \times 18.0\text{m} \times 3.00\text{m}$ high, and that all air is replaced in 30.0 min. Calculate the heat transfer per unit time in watts needed to warm the incoming cold air by 10.0°C , thus replacing the heat transferred by convection alone.

Strategy

Heat is used to raise the temperature of air so that $Q = mc\Delta T$. The rate of heat transfer is then Q/t , where t is the time for air turnover. We are given that ΔT is 10.0°C , but we must still find values for the mass of air and its

specific heat before we can calculate Q . The specific heat of air is a weighted average of the specific heats of nitrogen and oxygen, which gives $c = c_p \cong 1000 \text{ J/kg} \cdot ^\circ \text{C}$ from [\[link\]](#) (note that the specific heat at constant pressure must be used for this process).

Solution

1. Determine the mass of air from its density and the given volume of the house. The density is given from the density ρ and the volume

Equation:

$$m = \rho V = (1.29 \text{ kg/m}^3)(12.0 \text{ m} \times 18.0 \text{ m} \times 3.00 \text{ m}) = 836 \text{ kg}.$$

2. Calculate the heat transferred from the change in air temperature:

$$Q = mc\Delta T \text{ so that}$$

Equation:

$$Q = (836 \text{ kg})(1000 \text{ J/kg} \cdot ^\circ \text{C})(10.0^\circ \text{C}) = 8.36 \times 10^6 \text{ J}.$$

3. Calculate the heat transfer from the heat Q and the turnover time t . Since air is turned over in $t = 0.500 \text{ h} = 1800 \text{ s}$, the heat transferred per unit time is

Equation:

$$\frac{Q}{t} = \frac{8.36 \times 10^6 \text{ J}}{1800 \text{ s}} = 4.64 \text{ kW}.$$

Discussion

This rate of heat transfer is equal to the power consumed by about forty-six 100-W light bulbs. Newly constructed homes are designed for a turnover time of 2 hours or more, rather than 30 minutes for the house of this example. Weather stripping, caulking, and improved window seals are commonly employed. More extreme measures are sometimes taken in very cold (or hot) climates to achieve a tight standard of more than 6 hours for one air turnover. Still longer turnover times are unhealthy, because a minimum amount of fresh air is necessary to supply oxygen for breathing and to dilute household pollutants. The term used for the process by which

outside air leaks into the house from cracks around windows, doors, and the foundation is called “air infiltration.”

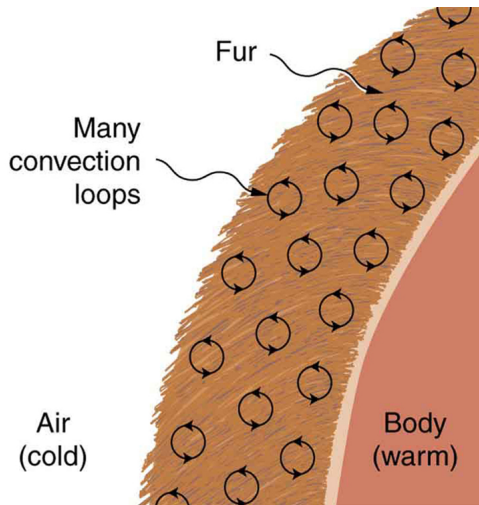
A cold wind is much more chilling than still cold air, because convection combines with conduction in the body to increase the rate at which energy is transferred away from the body. The table below gives approximate wind-chill factors, which are the temperatures of still air that produce the same rate of cooling as air of a given temperature and speed. Wind-chill factors are a dramatic reminder of convection’s ability to transfer heat faster than conduction. For example, a 15.0 m/s wind at 0°C has the chilling equivalent of still air at about -18°C .

Moving air temperature	Wind speed (m/s)				
(°C)	2	5	10	15	20
5	3	−1	−8	−10	−12
2	0	−7	−12	−16	−18
0	−2	−9	−15	−18	−20

Moving air temperature	Wind speed (m/s)				
−5	−7	−15	−22	−26	−29
−10	−12	−21	−29	−34	−36
−20	−23	−34	−44	−50	−52
−40	−44	−59	−73	−82	−84

Wind-Chill Factors

Although air can transfer heat rapidly by convection, it is a poor conductor and thus a good insulator. The amount of available space for airflow determines whether air acts as an insulator or conductor. The space between the inside and outside walls of a house, for example, is about 9 cm (3.5 in)—large enough for convection to work effectively. The addition of wall insulation prevents airflow, so heat loss (or gain) is decreased. Similarly, the gap between the two panes of a double-paned window is about 1 cm, which prevents convection and takes advantage of air's low conductivity to prevent greater loss. Fur, fiber, and fiberglass also take advantage of the low conductivity of air by trapping it in spaces too small to support convection, as shown in the figure. Fur and feathers are lightweight and thus ideal for the protection of animals.



Fur is filled with air, breaking it up into many small pockets. Convection is very slow here, because the loops are so small. The low conductivity of air makes fur a very good lightweight insulator.

Some interesting phenomena happen *when convection is accompanied by a phase change*. It allows us to cool off by sweating, even if the temperature of the surrounding air exceeds body temperature. Heat from the skin is required for sweat to evaporate from the skin, but without air flow, the air becomes saturated and evaporation stops. Air flow caused by convection replaces the saturated air by dry air and evaporation continues.

Example:

Calculate the Flow of Mass during Convection: Sweat-Heat Transfer away from the Body

The average person produces heat at the rate of about 120 W when at rest. At what rate must water evaporate from the body to get rid of all this energy? (This evaporation might occur when a person is sitting in the shade and surrounding temperatures are the same as skin temperature, eliminating heat transfer by other methods.)

Strategy

Energy is needed for a phase change ($Q = mL_v$). Thus, the energy loss per unit time is

Equation:

$$\frac{Q}{t} = \frac{mL_v}{t} = 120 \text{ W} = 120 \text{ J/s}.$$

We divide both sides of the equation by L_v to find that the mass evaporated per unit time is

Equation:

$$\frac{m}{t} = \frac{120 \text{ J/s}}{L_v}.$$

Solution

(1) Insert the value of the latent heat from [\[link\]](#), $L_v = 2430 \text{ kJ/kg} = 2430 \text{ J/g}$. This yields

Equation:

$$\frac{m}{t} = \frac{120 \text{ J/s}}{2430 \text{ J/g}} = 0.0494 \text{ g/s} = 2.96 \text{ g/min}.$$

Discussion

Evaporating about 3 g/min seems reasonable. This would be about 180 g (about 7 oz) per hour. If the air is very dry, the sweat may evaporate without even being noticed. A significant amount of evaporation also takes place in the lungs and breathing passages.

Another important example of the combination of phase change and convection occurs when water evaporates from the oceans. Heat is removed

from the ocean when water evaporates. If the water vapor condenses in liquid droplets as clouds form, heat is released in the atmosphere. Thus, there is an overall transfer of heat from the ocean to the atmosphere. This process is the driving power behind thunderheads, those great cumulus clouds that rise as much as 20.0 km into the stratosphere. Water vapor carried in by convection condenses, releasing tremendous amounts of energy. This energy causes the air to expand and rise, where it is colder. More condensation occurs in these colder regions, which in turn drives the cloud even higher. Such a mechanism is called positive feedback, since the process reinforces and accelerates itself. These systems sometimes produce violent storms, with lightning and hail, and constitute the mechanism driving hurricanes.



Cumulus clouds
are caused by
water vapor that
rises because of
convection. The
rise of clouds is
driven by a
positive
feedback
mechanism.
(credit: Mike
Love)



Convection
accompanied by a
phase change
releases the energy
needed to drive this
thunderhead into the
stratosphere. (credit:
Gerardo García
Moretti)



The phase change that
occurs when this
iceberg melts involves
tremendous heat

transfer. (credit:
Dominic Alves)

The movement of icebergs is another example of convection accompanied by a phase change. Suppose an iceberg drifts from Greenland into warmer Atlantic waters. Heat is removed from the warm ocean water when the ice melts and heat is released to the land mass when the iceberg forms on Greenland.

Exercise:

Check Your Understanding

Problem: Explain why using a fan in the summer feels refreshing!

Solution:

Using a fan increases the flow of air: warm air near your body is replaced by cooler air from elsewhere. Convection increases the rate of heat transfer so that moving air “feels” cooler than still air.

Summary

- Convection is heat transfer by the macroscopic movement of mass. Convection can be natural or forced and generally transfers thermal energy faster than conduction. [\[link\]](#) gives wind-chill factors, indicating that moving air has the same chilling effect of much colder stationary air. *Convection that occurs along with a phase change* can transfer energy from cold regions to warm ones.

Conceptual Questions

Exercise:

Problem:

One way to make a fireplace more energy efficient is to have an external air supply for the combustion of its fuel. Another is to have room air circulate around the outside of the fire box and back into the room. Detail the methods of heat transfer involved in each.

Exercise:**Problem:**

On cold, clear nights horses will sleep under the cover of large trees. How does this help them keep warm?

Problems & Exercises**Exercise:****Problem:**

At what wind speed does -10°C air cause the same chill factor as still air at -29°C ?

Solution:

10 m/s

Exercise:**Problem:**

At what temperature does still air cause the same chill factor as -5°C air moving at 15 m/s?

Exercise:

Problem:

The “steam” above a freshly made cup of instant coffee is really water vapor droplets condensing after evaporating from the hot coffee. What is the final temperature of 250 g of hot coffee initially at 90.0°C if 2.00 g evaporates from it? The coffee is in a Styrofoam cup, so other methods of heat transfer can be neglected.

Solution:

85.7°C

Exercise:**Problem:**

(a) How many kilograms of water must evaporate from a 60.0-kg woman to lower her body temperature by 0.750°C ?

(b) Is this a reasonable amount of water to evaporate in the form of perspiration, assuming the relative humidity of the surrounding air is low?

Exercise:**Problem:**

On a hot dry day, evaporation from a lake has just enough heat transfer to balance the $1.00\text{ kW}/\text{m}^2$ of incoming heat from the Sun. What mass of water evaporates in 1.00 h from each square meter? Explicitly show how you follow the steps in the [Problem-Solving Strategies for the Effects of Heat Transfer](#).

Solution:

1.48 kg

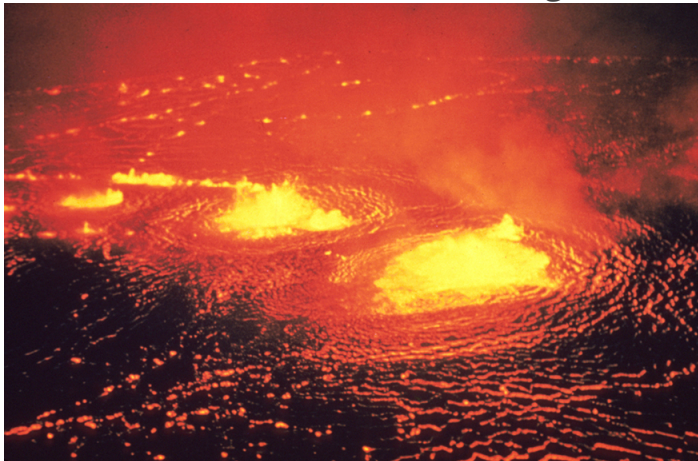
Exercise:

Problem:

One winter day, the climate control system of a large university classroom building malfunctions. As a result, 500 m^3 of excess cold air is brought in each minute. At what rate in kilowatts must heat transfer occur to warm this air by 10.0°C (that is, to bring the air to room temperature)?

Exercise:**Problem:**

The Kilauea volcano in Hawaii is the world's most active, disgorging about $5 \times 10^5 \text{ m}^3$ of 1200°C lava per day. What is the rate of heat transfer out of Earth by convection if this lava has a density of 2700 kg/m^3 and eventually cools to 30°C ? Assume that the specific heat of lava is the same as that of granite.



Lava flow on Kilauea volcano in Hawaii. (credit: J. P. Eaton, U.S. Geological Survey)

Solution:

$$2 \times 10^4 \text{ MW}$$

Exercise:**Problem:**

During heavy exercise, the body pumps 2.00 L of blood per minute to the surface, where it is cooled by 2.00°C . What is the rate of heat transfer from this forced convection alone, assuming blood has the same specific heat as water and its density is 1050 kg/m^3 ?

Exercise:**Problem:**

A person inhales and exhales 2.00 L of 37.0°C air, evaporating $4.00 \times 10^{-2}\text{ g}$ of water from the lungs and breathing passages with each breath.

- (a) How much heat transfer occurs due to evaporation in each breath?
- (b) What is the rate of heat transfer in watts if the person is breathing at a moderate rate of 18.0 breaths per minute?
- (c) If the inhaled air had a temperature of 20.0°C , what is the rate of heat transfer for warming the air?
- (d) Discuss the total rate of heat transfer as it relates to typical metabolic rates. Will this breathing be a major form of heat transfer for this person?

Solution:

- (a) 97.2 J
- (b) 29.2 W
- (c) 9.49 W
- (d) The total rate of heat loss would be $29.2\text{ W} + 9.49\text{ W} = 38.7\text{ W}$. While sleeping, our body consumes 83 W of power, while sitting it

consumes 120 to 210 W. Therefore, the total rate of heat loss from breathing will not be a major form of heat loss for this person.

Exercise:

Problem:

A glass coffee pot has a circular bottom with a 9.00-cm diameter in contact with a heating element that keeps the coffee warm with a continuous heat transfer rate of 50.0 W

(a) What is the temperature of the bottom of the pot, if it is 3.00 mm thick and the inside temperature is 60.0°C?

(b) If the temperature of the coffee remains constant and all of the heat transfer is removed by evaporation, how many grams per minute evaporate? Take the heat of vaporization to be 2340 kJ/kg.

Radiation

- Discuss heat transfer by radiation.
- Explain the power of different materials.

You can feel the heat transfer from a fire and from the Sun. Similarly, you can sometimes tell that the oven is hot without touching its door or looking inside—it may just warm you as you walk by. The space between the Earth and the Sun is largely empty, without any possibility of heat transfer by convection or conduction. In these examples, heat is transferred by radiation. That is, the hot body emits electromagnetic waves that are absorbed by our skin: no medium is required for electromagnetic waves to propagate. Different names are used for electromagnetic waves of different wavelengths: radio waves, microwaves, infrared **radiation**, visible light, ultraviolet radiation, X-rays, and gamma rays.

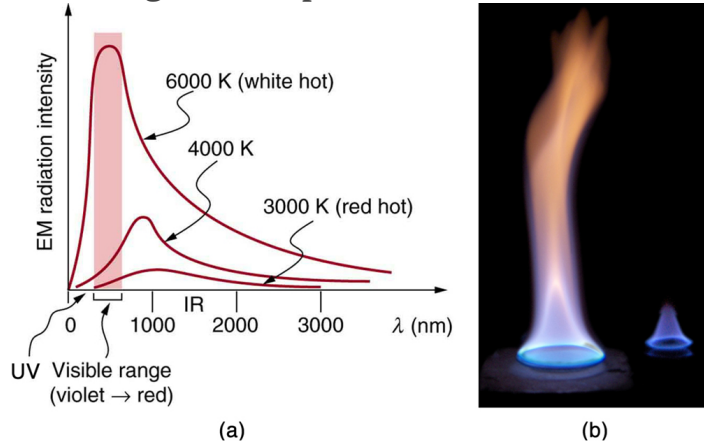


Most of the heat transfer from this fire to the observers is through infrared radiation. The visible light, although dramatic, transfers relatively little thermal energy. Convection transfers energy away from the observers as hot air rises, while conduction is negligibly slow here. Skin is very sensitive to infrared radiation, so that you can sense the presence of a fire

without looking at it directly. (credit:
Daniel X. O'Neil)

The energy of electromagnetic radiation depends on the wavelength (color) and varies over a wide range: a smaller wavelength (or higher frequency) corresponds to a higher energy. Because more heat is radiated at higher temperatures, a temperature change is accompanied by a color change. Take, for example, an electrical element on a stove, which glows from red to orange, while the higher-temperature steel in a blast furnace glows from yellow to white. The radiation you feel is mostly infrared, which corresponds to a lower temperature than that of the electrical element and the steel. The radiated energy depends on its intensity, which is represented in the figure below by the height of the distribution.

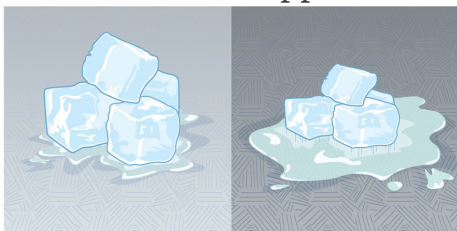
[Electromagnetic Waves](#) explains more about the electromagnetic spectrum and [Introduction to Quantum Physics](#) discusses how the decrease in wavelength corresponds to an increase in energy.



(a) A graph of the spectra of electromagnetic waves emitted from an ideal radiator at three different temperatures. The intensity or rate of radiation emission increases dramatically with temperature, and the spectrum shifts toward the visible and ultraviolet parts of the spectrum. The

shaded portion denotes the visible part of the spectrum. It is apparent that the shift toward the ultraviolet with temperature makes the visible appearance shift from red to white to blue as temperature increases. (b)
Note the variations in color corresponding to variations in flame temperature. (credit: Tuohirulla)

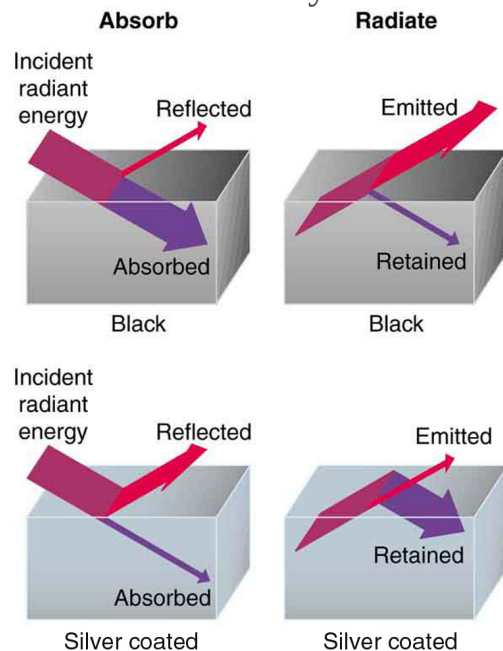
All objects absorb and emit electromagnetic radiation. The rate of heat transfer by radiation is largely determined by the color of the object. Black is the most effective, and white is the least effective. People living in hot climates generally avoid wearing black clothing, for instance (see [\[link\]](#)). Similarly, black asphalt in a parking lot will be hotter than adjacent gray sidewalk on a summer day, because black absorbs better than gray. The reverse is also true—black radiates better than gray. Thus, on a clear summer night, the asphalt will be colder than the gray sidewalk, because black radiates the energy more rapidly than gray. An *ideal radiator* is the same color as an *ideal absorber*, and captures all the radiation that falls on it. In contrast, white is a poor absorber and is also a poor radiator. A white object reflects all radiation, like a mirror. (A perfect, polished white surface is mirror-like in appearance, and a crushed mirror looks white.)



This illustration shows that the darker pavement is hotter than the lighter pavement (much more of the ice on the right has

melted), although both have been in the sunlight for the same time. The thermal conductivities of the pavements are the same.

Gray objects have a uniform ability to absorb all parts of the electromagnetic spectrum. Colored objects behave in similar but more complex ways, which gives them a particular color in the visible range and may make them special in other ranges of the nonvisible spectrum. Take, for example, the strong absorption of infrared radiation by the skin, which allows us to be very sensitive to it.



A black object is a good absorber and a good radiator, while a white (or silver) object is a poor absorber and a poor radiator. It is as if

radiation from the inside is reflected back into the silver object, whereas radiation from the inside of the black object is “absorbed” when it hits the surface and finds itself on the outside and is strongly emitted.

The rate of heat transfer by emitted radiation is determined by the **Stefan-Boltzmann law of radiation**:

Equation:

$$\frac{Q}{t} = \sigma e A T^4,$$

where $\sigma = 5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4$ is the Stefan-Boltzmann constant, A is the surface area of the object, and T is its absolute temperature in kelvin. The symbol e stands for the **emissivity** of the object, which is a measure of how well it radiates. An ideal jet-black (or black body) radiator has $e = 1$, whereas a perfect reflector has $e = 0$. Real objects fall between these two values. Take, for example, tungsten light bulb filaments which have an e of about 0.5, and carbon black (a material used in printer toner), which has the (greatest known) emissivity of about 0.99.

The radiation rate is directly proportional to the *fourth power* of the absolute temperature—a remarkably strong temperature dependence. Furthermore, the radiated heat is proportional to the surface area of the object. If you knock apart the coals of a fire, there is a noticeable increase in radiation due to an increase in radiating surface area.



A thermograph of part of a building shows temperature variations, indicating where heat transfer to the outside is most severe. Windows are a major region of heat transfer to the outside of homes. (credit: U.S. Army)

Skin is a remarkably good absorber and emitter of infrared radiation, having an emissivity of 0.97 in the infrared spectrum. Thus, we are all nearly (jet) black in the infrared, in spite of the obvious variations in skin color. This high infrared emissivity is why we can so easily feel radiation on our skin. It is also the basis for the use of night scopes used by law enforcement and the military to detect human beings. Even small temperature variations can be detected because of the T^4 dependence. Images, called *thermographs*, can be used medically to detect regions of abnormally high temperature in the body, perhaps indicative of disease. Similar techniques can be used to detect heat leaks in homes [\[link\]](#), optimize performance of blast furnaces, improve comfort levels in work environments, and even remotely map the Earth's temperature profile.

All objects emit and absorb radiation. The *net* rate of heat transfer by radiation (absorption minus emission) is related to both the temperature of the object and the temperature of its surroundings. Assuming that an object

with a temperature T_1 is surrounded by an environment with uniform temperature T_2 , the **net rate of heat transfer by radiation** is

Equation:

$$\frac{Q_{\text{net}}}{t} = \sigma e A (T_2^4 - T_1^4),$$

where e is the emissivity of the object alone. In other words, it does not matter whether the surroundings are white, gray, or black; the balance of radiation into and out of the object depends on how well it emits and absorbs radiation. When $T_2 > T_1$, the quantity Q_{net}/t is positive; that is, the net heat transfer is from hot to cold.

Note:

Take-Home Experiment: Temperature in the Sun

Place a thermometer out in the sunshine and shield it from direct sunlight using an aluminum foil. What is the reading? Now remove the shield, and note what the thermometer reads. Take a handkerchief soaked in nail polish remover, wrap it around the thermometer and place it in the sunshine. What does the thermometer read?

Example:

Calculate the Net Heat Transfer of a Person: Heat Transfer by Radiation

What is the rate of heat transfer by radiation, with an unclothed person standing in a dark room whose ambient temperature is 22.0°C . The person has a normal skin temperature of 33.0°C and a surface area of 1.50 m^2 . The emissivity of skin is 0.97 in the infrared, where the radiation takes place.

Strategy

We can solve this by using the equation for the rate of radiative heat transfer.

Solution

Insert the temperatures values $T_2 = 295 \text{ K}$ and $T_1 = 306 \text{ K}$, so that

Equation:

$$\frac{Q}{t} = \sigma e A (T_2^4 - T_1^4)$$

Equation:

$$= (5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4)(0.97)(1.50 \text{ m}^2) \left[(295 \text{ K})^4 - (306 \text{ K})^4 \right]$$

Equation:

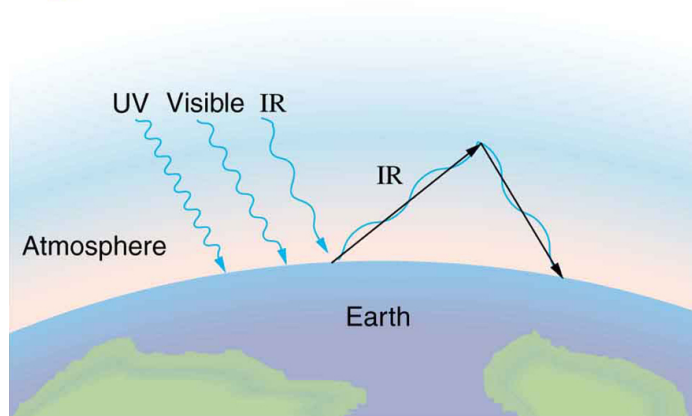
$$= -99 \text{ J/s} = -99 \text{ W}.$$

Discussion

This value is a significant rate of heat transfer to the environment (note the minus sign), considering that a person at rest may produce energy at the rate of 125 W and that conduction and convection will also be transferring energy to the environment. Indeed, we would probably expect this person to feel cold. Clothing significantly reduces heat transfer to the environment by many methods, because clothing slows down both conduction and convection, and has a lower emissivity (especially if it is white) than skin.

The Earth receives almost all its energy from radiation of the Sun and reflects some of it back into outer space. Because the Sun is hotter than the Earth, the net energy flux is from the Sun to the Earth. However, the rate of energy transfer is less than the equation for the radiative heat transfer would predict because the Sun does not fill the sky. The average emissivity (e) of the Earth is about 0.65, but the calculation of this value is complicated by the fact that the highly reflective cloud coverage varies greatly from day to day. There is a negative feedback (one in which a change produces an effect that opposes that change) between clouds and heat transfer; greater temperatures evaporate more water to form more clouds, which reflect more radiation back into space, reducing the temperature. The often mentioned **greenhouse effect** is directly related to the variation of the Earth's emissivity with radiation type (see the figure given below). The greenhouse

effect is a natural phenomenon responsible for providing temperatures suitable for life on Earth. The Earth's relatively constant temperature is a result of the energy balance between the incoming solar radiation and the energy radiated from the Earth. Most of the infrared radiation emitted from the Earth is absorbed by carbon dioxide (CO_2) and water (H_2O) in the atmosphere and then re-radiated back to the Earth or into outer space. Re-radiation back to the Earth maintains its surface temperature about 40°C higher than it would be if there was no atmosphere, similar to the way glass increases temperatures in a greenhouse.

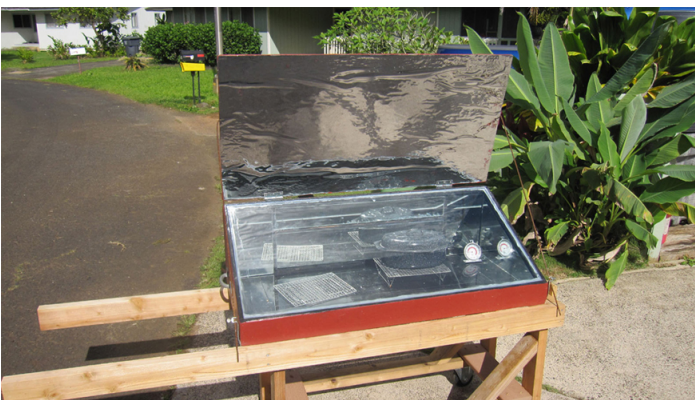


The greenhouse effect is a name given to the trapping of energy in the Earth's atmosphere by a process similar to that used in greenhouses. The atmosphere, like window glass, is transparent to incoming visible radiation and most of the Sun's infrared. These wavelengths are absorbed by the Earth and re-emitted as infrared. Since Earth's temperature is much lower than that of the Sun, the infrared radiated by the Earth has a much longer wavelength. The atmosphere, like glass, traps these longer infrared rays, keeping the Earth warmer than it would otherwise

be. The amount of trapping depends on concentrations of trace gases like carbon dioxide, and a change in the concentration of these gases is believed to affect the Earth's surface temperature.

The greenhouse effect is also central to the discussion of global warming due to emission of carbon dioxide and methane (and other so-called greenhouse gases) into the Earth's atmosphere from industrial production and farming. Changes in global climate could lead to more intense storms, precipitation changes (affecting agriculture), reduction in rain forest biodiversity, and rising sea levels.

Heating and cooling are often significant contributors to energy use in individual homes. Current research efforts into developing environmentally friendly homes quite often focus on reducing conventional heating and cooling through better building materials, strategically positioning windows to optimize radiation gain from the Sun, and opening spaces to allow convection. It is possible to build a zero-energy house that allows for comfortable living in most parts of the United States with hot and humid summers and cold winters.



This simple but effective solar cooker uses the greenhouse effect and reflective material to trap and retain solar energy. Made of inexpensive,

durable materials, it saves money and labor, and is of particular economic value in energy-poor developing countries. (credit: E.B. Kauai)

Conversely, dark space is very cold, about 3K (-454°F), so that the Earth radiates energy into the dark sky. Owing to the fact that clouds have lower emissivity than either oceans or land masses, they reflect some of the radiation back to the surface, greatly reducing heat transfer into dark space, just as they greatly reduce heat transfer into the atmosphere during the day. The rate of heat transfer from soil and grasses can be so rapid that frost may occur on clear summer evenings, even in warm latitudes.

Exercise:

Check Your Understanding

Problem:

What is the change in the rate of the radiated heat by a body at the temperature $T_1 = 20^{\circ}\text{C}$ compared to when the body is at the temperature $T_2 = 40^{\circ}\text{C}$?

Solution:

The radiated heat is proportional to the fourth power of the *absolute temperature*. Because $T_1 = 293\text{ K}$ and $T_2 = 313\text{ K}$, the rate of heat transfer increases by about 30 percent of the original rate.

Note:

Career Connection: Energy Conservation Consultation

The cost of energy is generally believed to remain very high for the foreseeable future. Thus, passive control of heat loss in both commercial and domestic housing will become increasingly important. Energy consultants measure and analyze the flow of energy into and out of houses

and ensure that a healthy exchange of air is maintained inside the house. The job prospects for an energy consultant are strong.

Note:

Problem-Solving Strategies for the Methods of Heat Transfer

1. *Examine the situation to determine what type of heat transfer is involved.*
2. *Identify the type(s) of heat transfer—conduction, convection, or radiation.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is very useful.*
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).*
6. For conduction, equation $\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d}$ is appropriate. [\[link\]](#) lists thermal conductivities. For convection, determine the amount of matter moved and use equation $Q = mc\Delta T$, to calculate the heat transfer involved in the temperature change of the fluid. If a phase change accompanies convection, equation $Q = mL_f$ or $Q = mL_v$ is appropriate to find the heat transfer involved in the phase change. [\[link\]](#) lists information relevant to phase change. For radiation, equation $\frac{Q_{\text{net}}}{t} = \sigma eA(T_2^4 - T_1^4)$ gives the net heat transfer rate.
7. *Insert the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units.*
8. *Check the answer to see if it is reasonable. Does it make sense?*

Summary

- Radiation is the rate of heat transfer through the emission or absorption of electromagnetic waves.

- The rate of heat transfer depends on the surface area and the fourth power of the absolute temperature:

Equation:

$$\frac{Q}{t} = \sigma e A T^4,$$

where $\sigma = 5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4$ is the Stefan-Boltzmann constant and e is the emissivity of the body. For a black body, $e = 1$ whereas a shiny white or perfect reflector has $e = 0$, with real objects having values of e between 1 and 0. The net rate of heat transfer by radiation is

Equation:

$$\frac{Q_{\text{net}}}{t} = \sigma e A (T_2^4 - T_1^4)$$

where T_1 is the temperature of an object surrounded by an environment with uniform temperature T_2 and e is the emissivity of the *object*.

Conceptual Questions

Exercise:

Problem:

When watching a daytime circus in a large, dark-colored tent, you sense significant heat transfer from the tent. Explain why this occurs.

Exercise:

Problem:

Satellites designed to observe the radiation from cold (3 K) dark space have sensors that are shaded from the Sun, Earth, and Moon and that are cooled to very low temperatures. Why must the sensors be at low temperature?

Exercise:

Problem: Why are cloudy nights generally warmer than clear ones?

Exercise:

Problem:

Why are thermometers that are used in weather stations shielded from the sunshine? What does a thermometer measure if it is shielded from the sunshine and also if it is not?

Exercise:

Problem:

On average, would Earth be warmer or cooler without the atmosphere? Explain your answer.

Problems & Exercises

Exercise:

Problem:

At what net rate does heat radiate from a 275-m^2 black roof on a night when the roof's temperature is 30.0°C and the surrounding temperature is 15.0°C ? The emissivity of the roof is 0.900.

Solution:

-21.7 kW

Note that the negative answer implies heat loss to the surroundings.

Exercise:

Problem:

(a) Cherry-red embers in a fireplace are at 850°C and have an exposed area of 0.200 m^2 and an emissivity of 0.980. The surrounding room has a temperature of 18.0°C . If 50% of the radiant energy enters the room, what is the net rate of radiant heat transfer in kilowatts? (b) Does your answer support the contention that most of the heat transfer into a room by a fireplace comes from infrared radiation?

Exercise:**Problem:**

Radiation makes it impossible to stand close to a hot lava flow. Calculate the rate of heat transfer by radiation from 1.00 m^2 of 1200°C fresh lava into 30.0°C surroundings, assuming lava's emissivity is 1.00.

Solution:

−266 kW

Exercise:**Problem:**

(a) Calculate the rate of heat transfer by radiation from a car radiator at 110°C into a 50.0°C environment, if the radiator has an emissivity of 0.750 and a 1.20-m^2 surface area. (b) Is this a significant fraction of the heat transfer by an automobile engine? To answer this, assume a horsepower of 200 hp (1.5 kW) and the efficiency of automobile engines as 25%.

Exercise:

Problem:

Find the net rate of heat transfer by radiation from a skier standing in the shade, given the following. She is completely clothed in white (head to foot, including a ski mask), the clothes have an emissivity of 0.200 and a surface temperature of 10.0°C , the surroundings are at -15.0°C , and her surface area is 1.60 m^2 .

Solution:

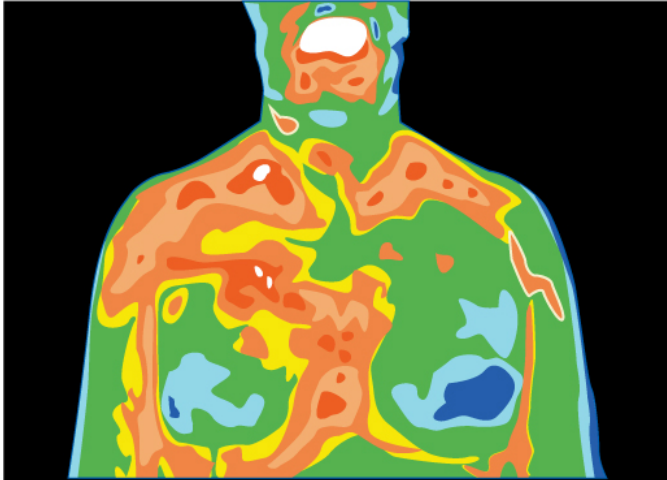
-36.0 W

Exercise:**Problem:**

Suppose you walk into a sauna that has an ambient temperature of 50.0°C . (a) Calculate the rate of heat transfer to you by radiation given your skin temperature is 37.0°C , the emissivity of skin is 0.98, and the surface area of your body is 1.50 m^2 . (b) If all other forms of heat transfer are balanced (the net heat transfer is zero), at what rate will your body temperature increase if your mass is 75.0 kg ?

Exercise:**Problem:**

Thermography is a technique for measuring radiant heat and detecting variations in surface temperatures that may be medically, environmentally, or militarily meaningful. (a) What is the percent increase in the rate of heat transfer by radiation from a given area at a temperature of 34.0°C compared with that at 33.0°C , such as on a person's skin? (b) What is the percent increase in the rate of heat transfer by radiation from a given area at a temperature of 34.0°C compared with that at 20.0°C , such as for warm and cool automobile hoods?



Artist's rendition of a thermograph of a patient's upper body, showing the distribution of heat represented by different colors.

Solution:

(a) 1.31%

(b) 20.5%

Exercise:

Problem:

The Sun radiates like a perfect black body with an emissivity of exactly 1. (a) Calculate the surface temperature of the Sun, given that it is a sphere with a 7.00×10^8 -m radius that radiates 3.80×10^{26} W into 3-K space. (b) How much power does the Sun radiate per square meter of its surface? (c) How much power in watts per square meter is that value at the distance of Earth, 1.50×10^{11} m away? (This number is called the solar constant.)

Exercise:

Problem:

A large body of lava from a volcano has stopped flowing and is slowly cooling. The interior of the lava is at 1200°C , its surface is at 450°C , and the surroundings are at 27.0°C . (a) Calculate the rate at which energy is transferred by radiation from 1.00 m^2 of surface lava into the surroundings, assuming the emissivity is 1.00. (b) Suppose heat conduction to the surface occurs at the same rate. What is the thickness of the lava between the 450°C surface and the 1200°C interior, assuming that the lava's conductivity is the same as that of brick?

Solution:

(a) -15.0 kW

(b) 4.2 cm

Exercise:**Problem:**

Calculate the temperature the entire sky would have to be in order to transfer energy by radiation at 1000 W/m^2 —about the rate at which the Sun radiates when it is directly overhead on a clear day. This value is the effective temperature of the sky, a kind of average that takes account of the fact that the Sun occupies only a small part of the sky but is much hotter than the rest. Assume that the body receiving the energy has a temperature of 27.0°C .

Exercise:

Problem:

(a) A shirtless rider under a circus tent feels the heat radiating from the sunlit portion of the tent. Calculate the temperature of the tent canvas based on the following information: The shirtless rider's skin temperature is 34.0°C and has an emissivity of 0.970. The exposed area of skin is 0.400 m^2 . He receives radiation at the rate of 20.0 W —half what you would calculate if the entire region behind him was hot. The rest of the surroundings are at 34.0°C . (b) Discuss how this situation would change if the sunlit side of the tent was nearly pure white and if the rider was covered by a white tunic.

Solution:

(a) 48.5°C

(b) A pure white object reflects more of the radiant energy that hits it, so a white tent would prevent more of the sunlight from heating up the inside of the tent, and the white tunic would prevent that heat which entered the tent from heating the rider. Therefore, with a white tent, the temperature would be lower than 48.5°C , and the rate of radiant heat transferred to the rider would be less than 20.0 W .

Exercise:**Problem: Integrated Concepts**

One 30.0°C day the relative humidity is 75.0%, and that evening the temperature drops to 20.0°C , well below the dew point. (a) How many grams of water condense from each cubic meter of air? (b) How much heat transfer occurs by this condensation? (c) What temperature increase could this cause in dry air?

Exercise:**Problem: Integrated Concepts**

Large meteors sometimes strike the Earth, converting most of their kinetic energy into thermal energy. (a) What is the kinetic energy of a 10^9 kg meteor moving at 25.0 km/s? (b) If this meteor lands in a deep ocean and 80% of its kinetic energy goes into heating water, how many kilograms of water could it raise by 5.0°C ? (c) Discuss how the energy of the meteor is more likely to be deposited in the ocean and the likely effects of that energy.

Solution:

(a) $3 \times 10^{17} \text{ J}$

(b) $1 \times 10^{13} \text{ kg}$

(c) When a large meteor hits the ocean, it causes great tidal waves, dissipating large amount of its energy in the form of kinetic energy of the water.

Exercise:

Problem: Integrated Concepts

Frozen waste from airplane toilets has sometimes been accidentally ejected at high altitude. Ordinarily it breaks up and disperses over a large area, but sometimes it holds together and strikes the ground. Calculate the mass of 0°C ice that can be melted by the conversion of kinetic and gravitational potential energy when a 20.0 kg piece of frozen waste is released at 12.0 km altitude while moving at 250 m/s and strikes the ground at 100 m/s (since less than 20.0 kg melts, a significant mess results).

Exercise:

Problem: Integrated Concepts

(a) A large electrical power facility produces 1600 MW of “waste heat,” which is dissipated to the environment in cooling towers by warming air flowing through the towers by 5.00°C . What is the

necessary flow rate of air in m^3/s ? (b) Is your result consistent with the large cooling towers used by many large electrical power plants?

Solution:

(a) $3.44 \times 10^5 \text{ m}^3/\text{s}$

(b) This is equivalent to 12 million cubic feet of air per second. That is tremendous. This is too large to be dissipated by heating the air by only 5°C . Many of these cooling towers use the circulation of cooler air over warmer water to increase the rate of evaporation. This would allow much smaller amounts of air necessary to remove such a large amount of heat because evaporation removes larger quantities of heat than was considered in part (a).

Exercise:

Problem: Integrated Concepts

(a) Suppose you start a workout on a Stairmaster, producing power at the same rate as climbing 116 stairs per minute. Assuming your mass is 76.0 kg and your efficiency is 20.0% , how long will it take for your body temperature to rise 1.00°C if all other forms of heat transfer in and out of your body are balanced? (b) Is this consistent with your experience in getting warm while exercising?

Exercise:

Problem: Integrated Concepts

A 76.0-kg person suffering from hypothermia comes indoors and shivers vigorously. How long does it take the heat transfer to increase the person's body temperature by 2.00°C if all other forms of heat transfer are balanced?

Solution:

20.9 min

Exercise:**Problem: Integrated Concepts**

In certain large geographic regions, the underlying rock is hot. Wells can be drilled and water circulated through the rock for heat transfer for the generation of electricity. (a) Calculate the heat transfer that can be extracted by cooling 1.00 km^3 of granite by 100°C . (b) How long will this take if heat is transferred at a rate of 300 MW, assuming no heat transfers back into the 1.00 km of rock by its surroundings?

Exercise:**Problem: Integrated Concepts**

Heat transfers from your lungs and breathing passages by evaporating water. (a) Calculate the maximum number of grams of water that can be evaporated when you inhale 1.50 L of 37°C air with an original relative humidity of 40.0%. (Assume that body temperature is also 37°C .) (b) How many joules of energy are required to evaporate this amount? (c) What is the rate of heat transfer in watts from this method, if you breathe at a normal resting rate of 10.0 breaths per minute?

Solution:

(a) $3.96 \times 10^{-2} \text{ g}$

(b) 96.2 J

(c) 16.0 W

Exercise:**Problem: Integrated Concepts**

(a) What is the temperature increase of water falling 55.0 m over Niagara Falls? (b) What fraction must evaporate to keep the temperature constant?

Exercise:**Problem: Integrated Concepts**

Hot air rises because it has expanded. It then displaces a greater volume of cold air, which increases the buoyant force on it. (a) Calculate the ratio of the buoyant force to the weight of 50.0°C air surrounded by 20.0°C air. (b) What energy is needed to cause 1.00 m^3 of air to go from 20.0°C to 50.0°C ? (c) What gravitational potential energy is gained by this volume of air if it rises 1.00 m ? Will this cause a significant cooling of the air?

Solution:

(a) 1.102

(b) $2.79 \times 10^4\text{ J}$

(c) 12.6 J. This will not cause a significant cooling of the air because it is much less than the energy found in part (b), which is the energy required to warm the air from 20.0°C to 50.0°C .

Exercise:**Problem: Unreasonable Results**

(a) What is the temperature increase of an 80.0 kg person who consumes 2500 kcal of food in one day with 95.0% of the energy transferred as heat to the body? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) 36°C

(b) Any temperature increase greater than about 3°C would be unreasonably large. In this case the final temperature of the person would rise to 73°C (163°F).

(c) The assumption of 95% heat retention is unreasonable.

Exercise:

Problem: Unreasonable Results

A slightly deranged Arctic inventor surrounded by ice thinks it would be much less mechanically complex to cool a car engine by melting ice on it than by having a water-cooled system with a radiator, water pump, antifreeze, and so on. (a) If 80.0% of the energy in 1.00 gal of gasoline is converted into “waste heat” in a car engine, how many kilograms of 0°C ice could it melt? (b) Is this a reasonable amount of ice to carry around to cool the engine for 1.00 gal of gasoline consumption? (c) What premises or assumptions are unreasonable?

Exercise:

Problem: Unreasonable Results

(a) Calculate the rate of heat transfer by conduction through a window with an area of 1.00 m^2 that is 0.750 cm thick, if its inner surface is at 22.0°C and its outer surface is at 35.0°C. (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) 1.46 kW

(b) Very high power loss through a window. An electric heater of this power can keep an entire room warm.

(c) The surface temperatures of the window do not differ by as great an amount as assumed. The inner surface will be warmer, and the outer surface will be cooler.

Exercise:

Problem: Unreasonable Results

A meteorite 1.20 cm in diameter is so hot immediately after penetrating the atmosphere that it radiates 20.0 kW of power. (a) What is its temperature, if the surroundings are at 20.0°C and it has an emissivity of 0.800? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Exercise:

Problem: Construct Your Own Problem

Consider a new model of commercial airplane having its brakes tested as a part of the initial flight permission procedure. The airplane is brought to takeoff speed and then stopped with the brakes alone. Construct a problem in which you calculate the temperature increase of the brakes during this process. You may assume most of the kinetic energy of the airplane is converted to thermal energy in the brakes and surrounding materials, and that little escapes. Note that the brakes are expected to become so hot in this procedure that they ignite and, in order to pass the test, the airplane must be able to withstand the fire for some time without a general conflagration.

Exercise:

Problem: Construct Your Own Problem

Consider a person outdoors on a cold night. Construct a problem in which you calculate the rate of heat transfer from the person by all three heat transfer methods. Make the initial circumstances such that at rest the person will have a net heat transfer and then decide how much physical activity of a chosen type is necessary to balance the rate of heat transfer. Among the things to consider are the size of the person, type of clothing, initial metabolic rate, sky conditions, amount of water evaporated, and volume of air breathed. Of course, there are many other factors to consider and your instructor may wish to guide you in the assumptions made as well as the detail of analysis and method of presenting your results.

Glossary

emissivity

measure of how well an object radiates

greenhouse effect

warming of the Earth that is due to gases such as carbon dioxide and methane that absorb infrared radiation from the Earth's surface and reradiate it in all directions, thus sending a fraction of it back toward the surface of the Earth

net rate of heat transfer by radiation

is $\frac{Q_{\text{net}}}{t} = \sigma e A (T_2^4 - T_1^4)$

radiation

energy transferred by electromagnetic waves directly as a result of a temperature difference

Stefan-Boltzmann law of radiation

$\frac{Q}{t} = \sigma e A T^4$ where σ is the Stefan-Boltzmann constant, A is the surface area of the object, T is the absolute temperature, and e is the emissivity

Introduction to Thermodynamics

class="introduction"

A steam
engine
uses heat
transfer
to do
work.

Tourists
regularly
ride this
narrow-
gauge
steam
engine
train near
the San
Juan
Skyway
in
Durango,
Colorado
, part of
the
National
Scenic
Byways
Program.
(credit:
Dennis
Adams)



Heat transfer is energy in transit, and it can be used to do work. It can also be converted to any other form of energy. A car engine, for example, burns fuel for heat transfer into a gas. Work is done by the gas as it exerts a force through a distance, converting its energy into a variety of other forms—into the car's kinetic or gravitational potential energy; into electrical energy to run the spark plugs, radio, and lights; and back into stored energy in the car's battery. But most of the heat transfer produced from burning fuel in the engine does not do work on the gas. Rather, the energy is released into the environment, implying that the engine is quite inefficient.

It is often said that modern gasoline engines cannot be made to be significantly more efficient. We hear the same about heat transfer to electrical energy in large power stations, whether they are coal, oil, natural gas, or nuclear powered. Why is that the case? Is the inefficiency caused by design problems that could be solved with better engineering and superior materials? Is it part of some money-making conspiracy by those who sell energy? Actually, the truth is more interesting, and reveals much about the nature of heat transfer.

Basic physical laws govern how heat transfer for doing work takes place and place insurmountable limits onto its efficiency. This chapter will explore these laws as well as many applications and concepts associated

with them. These topics are part of *thermodynamics*—the study of heat transfer and its relationship to doing work.

The First Law of Thermodynamics

- Define the first law of thermodynamics.
- Describe how conservation of energy relates to the first law of thermodynamics.
- Identify instances of the first law of thermodynamics working in everyday situations, including biological metabolism.
- Calculate changes in the internal energy of a system, after accounting for heat transfer and work done.



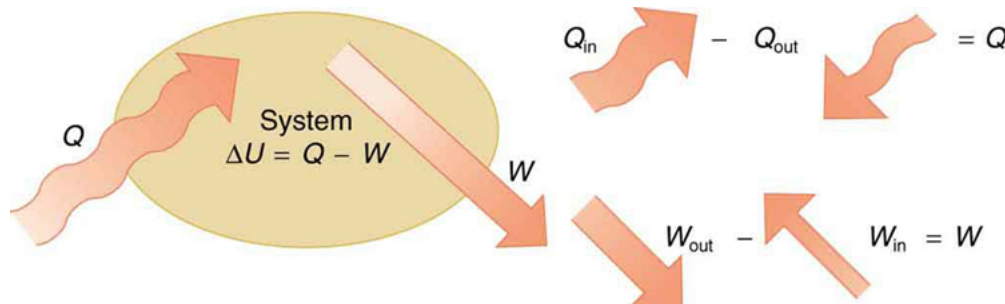
This boiling tea kettle represents energy in motion. The water in the kettle is turning to water vapor because heat is being transferred from the stove to the kettle. As the entire system gets hotter, work is done—from the evaporation of the water to the whistling of the kettle. (credit: Gina Hamilton)

If we are interested in how heat transfer is converted into doing work, then the conservation of energy principle is important. The first law of thermodynamics applies the conservation of energy principle to systems where heat transfer and doing work are the methods of transferring energy into and out of the system. The **first law of thermodynamics** states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system. In equation form, the first law of thermodynamics is

Equation:

$$\Delta U = Q - W.$$

Here ΔU is the *change in internal energy* U of the system. Q is the *net heat transferred into the system*—that is, Q is the sum of all heat transfer into and out of the system. W is the *net work done by the system*—that is, W is the sum of all work done on or by the system. We use the following sign conventions: if Q is positive, then there is a net heat transfer into the system; if W is positive, then there is net work done by the system. So positive Q adds energy to the system and positive W takes energy from the system. Thus $\Delta U = Q - W$. Note also that if more heat transfer into the system occurs than work done, the difference is stored as internal energy. Heat engines are a good example of this—heat transfer into them takes place so that they can do work. (See [\[link\]](#).) We will now examine Q , W , and ΔU further.



The first law of thermodynamics is the conservation-of-energy principle stated for a system where heat and work are the methods of transferring energy for a system in

thermal equilibrium. Q represents the net heat transfer—it is the sum of all heat transfers into and out of the system. Q is positive for net heat transfer *into* the system. W is the total work done on and by the system. W is positive when more work is done *by* the system than on it. The change in the internal energy of the system, ΔU , is related to heat and work by the first law of thermodynamics, $\Delta U = Q - W$.

Note:

Making Connections: Law of Thermodynamics and Law of Conservation of Energy

The first law of thermodynamics is actually the law of conservation of energy stated in a form most useful in thermodynamics. The first law gives the relationship between heat transfer, work done, and the change in internal energy of a system.

Heat Q and Work W

Heat transfer (Q) and doing work (W) are the two everyday means of bringing energy into or taking energy out of a system. The processes are quite different. Heat transfer, a less organized process, is driven by temperature differences. Work, a quite organized process, involves a macroscopic force exerted through a distance. Nevertheless, heat and work can produce identical results. For example, both can cause a temperature increase. Heat transfer into a system, such as when the Sun warms the air in a bicycle tire, can increase its temperature, and so can work done on the system, as when the bicyclist pumps air into the tire. Once the temperature increase has occurred, it is impossible to tell whether it was caused by heat transfer or by doing work. This uncertainty is an important point. Heat transfer and work are both energy in transit—neither is stored as such in a

system. However, both can change the internal energy U of a system. Internal energy is a form of energy completely different from either heat or work.

Internal Energy U

We can think about the internal energy of a system in two different but consistent ways. The first is the atomic and molecular view, which examines the system on the atomic and molecular scale. The **internal energy** U of a system is the sum of the kinetic and potential energies of its atoms and molecules. Recall that kinetic plus potential energy is called mechanical energy. Thus internal energy is the sum of atomic and molecular mechanical energy. Because it is impossible to keep track of all individual atoms and molecules, we must deal with averages and distributions. A second way to view the internal energy of a system is in terms of its macroscopic characteristics, which are very similar to atomic and molecular average values.

Macroscopically, we define the change in internal energy ΔU to be that given by the first law of thermodynamics:

Equation:

$$\Delta U = Q - W.$$

Many detailed experiments have verified that $\Delta U = Q - W$, where ΔU is the change in total kinetic and potential energy of all atoms and molecules in a system. It has also been determined experimentally that the internal energy U of a system depends only on the state of the system and *not how it reached that state*. More specifically, U is found to be a function of a few macroscopic quantities (pressure, volume, and temperature, for example), independent of past history such as whether there has been heat transfer or work done. This independence means that if we know the state of a system, we can calculate changes in its internal energy U from a few macroscopic variables.

Note:**Making Connections: Macroscopic and Microscopic**

In thermodynamics, we often use the macroscopic picture when making calculations of how a system behaves, while the atomic and molecular picture gives underlying explanations in terms of averages and distributions. We shall see this again in later sections of this chapter. For example, in the topic of entropy, calculations will be made using the atomic and molecular view.

To get a better idea of how to think about the internal energy of a system, let us examine a system going from State 1 to State 2. The system has internal energy U_1 in State 1, and it has internal energy U_2 in State 2, no matter how it got to either state. So the change in internal energy $\Delta U = U_2 - U_1$ is independent of what caused the change. In other words, ΔU is *independent of path*. By path, we mean the method of getting from the starting point to the ending point. Why is this independence important? Note that $\Delta U = Q - W$. Both Q and W depend on path, but ΔU does not. This path independence means that internal energy U is easier to consider than either heat transfer or work done.

Example:**Calculating Change in Internal Energy: The Same Change in U is Produced by Two Different Processes**

(a) Suppose there is heat transfer of 40.00 J to a system, while the system does 10.00 J of work. Later, there is heat transfer of 25.00 J out of the system while 4.00 J of work is done on the system. What is the net change in internal energy of the system?

(b) What is the change in internal energy of a system when a total of 150.00 J of heat transfer occurs out of (from) the system and 159.00 J of work is done on the system? (See [\[link\]](#)).

Strategy

In part (a), we must first find the net heat transfer and net work done from the given information. Then the first law of thermodynamics

($\Delta U = Q - W$) can be used to find the change in internal energy. In part (b), the net heat transfer and work done are given, so the equation can be used directly.

Solution for (a)

The net heat transfer is the heat transfer into the system minus the heat transfer out of the system, or

Equation:

$$Q = 40.00 \text{ J} - 25.00 \text{ J} = 15.00 \text{ J}.$$

Similarly, the total work is the work done by the system minus the work done on the system, or

Equation:

$$W = 10.00 \text{ J} - 4.00 \text{ J} = 6.00 \text{ J}.$$

Thus the change in internal energy is given by the first law of thermodynamics:

Equation:

$$\Delta U = Q - W = 15.00 \text{ J} - 6.00 \text{ J} = 9.00 \text{ J}.$$

We can also find the change in internal energy for each of the two steps. First, consider 40.00 J of heat transfer in and 10.00 J of work out, or

Equation:

$$\Delta U_1 = Q_1 - W_1 = 40.00 \text{ J} - 10.00 \text{ J} = 30.00 \text{ J}.$$

Now consider 25.00 J of heat transfer out and 4.00 J of work in, or

Equation:

$$\Delta U_2 = Q_2 - W_2 = -25.00 \text{ J} - (-4.00 \text{ J}) = -21.00 \text{ J}.$$

The total change is the sum of these two steps, or

Equation:

$$\Delta U = \Delta U_1 + \Delta U_2 = 30.00 \text{ J} + (-21.00 \text{ J}) = 9.00 \text{ J}.$$

Discussion on (a)

No matter whether you look at the overall process or break it into steps, the change in internal energy is the same.

Solution for (b)

Here the net heat transfer and total work are given directly to be

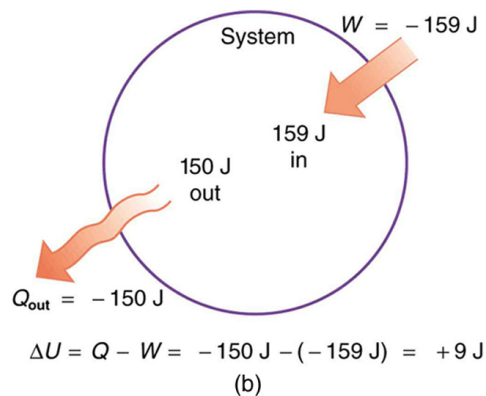
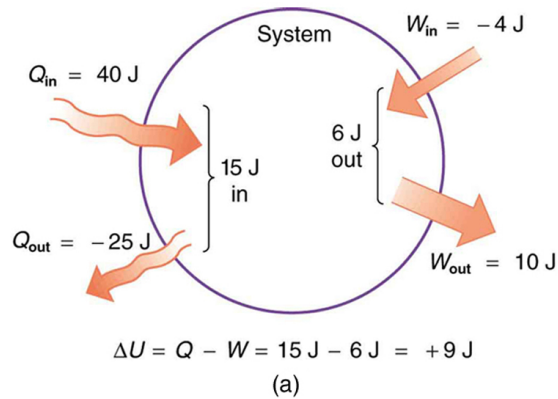
$Q = -150.00 \text{ J}$ and $W = -159.00 \text{ J}$, so that

Equation:

$$\Delta U = Q - W = -150.00 \text{ J} - (-159.00 \text{ J}) = 9.00 \text{ J}.$$

Discussion on (b)

A very different process in part (b) produces the same 9.00-J change in internal energy as in part (a). Note that the change in the system in both parts is related to ΔU and not to the individual Q s or W s involved. The system ends up in the *same* state in both (a) and (b). Parts (a) and (b) present two different paths for the system to follow between the same starting and ending points, and the change in internal energy for each is the same—it is independent of path.



Two different processes produce the same change in a system. (a) A total of 15.00 J of heat transfer occurs into the system, while work takes out a total of 6.00 J. The change in internal energy is $\Delta U = Q - W = 9.00 \text{ J}$. (b) Heat transfer removes 150.00 J from the system while work puts 159.00 J into it, producing an increase of 9.00 J in internal energy. If the system starts out in the same state in (a) and (b), it will end up in the same final state in either case—its final state is related to internal

energy, not how that energy
was acquired.

Human Metabolism and the First Law of Thermodynamics

Human metabolism is the conversion of food into heat transfer, work, and stored fat. Metabolism is an interesting example of the first law of thermodynamics in action. We now take another look at these topics via the first law of thermodynamics. Considering the body as the system of interest, we can use the first law to examine heat transfer, doing work, and internal energy in activities ranging from sleep to heavy exercise. What are some of the major characteristics of heat transfer, doing work, and energy in the body? For one, body temperature is normally kept constant by heat transfer to the surroundings. This means Q is negative. Another fact is that the body usually does work on the outside world. This means W is positive. In such situations, then, the body loses internal energy, since $\Delta U = Q - W$ is negative.

Now consider the effects of eating. Eating increases the internal energy of the body by adding chemical potential energy (this is an unromantic view of a good steak). The body *metabolizes* all the food we consume. Basically, metabolism is an oxidation process in which the chemical potential energy of food is released. This implies that food input is in the form of work. Food energy is reported in a special unit, known as the Calorie. This energy is measured by burning food in a calorimeter, which is how the units are determined.

In chemistry and biochemistry, one calorie (spelled with a *lowercase c*) is defined as the energy (or heat transfer) required to raise the temperature of one gram of pure water by one degree Celsius. Nutritionists and weight-watchers tend to use the *dietary* calorie, which is frequently called a Calorie (spelled with a *capital C*). One food Calorie is the energy needed to raise the temperature of one *kilogram* of water by one degree Celsius. This

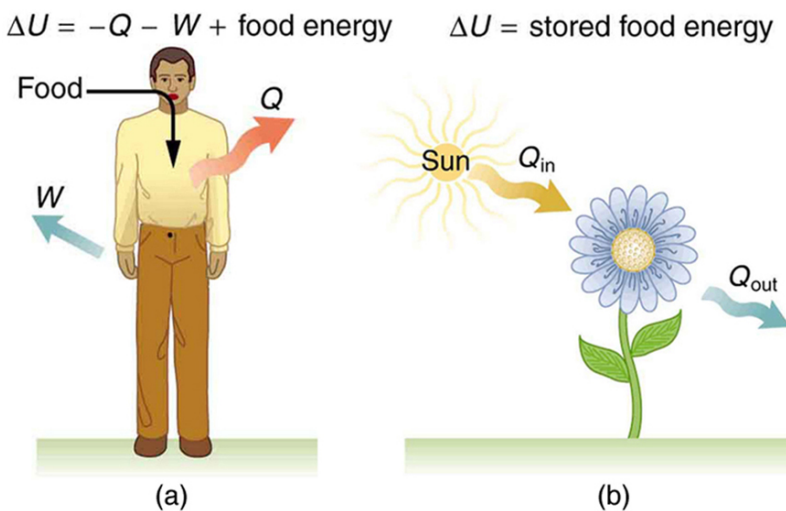
means that one dietary Calorie is equal to one kilocalorie for the chemist, and one must be careful to avoid confusion between the two.

Again, consider the internal energy the body has lost. There are three places this internal energy can go—to heat transfer, to doing work, and to stored fat (a tiny fraction also goes to cell repair and growth). Heat transfer and doing work take internal energy out of the body, and food puts it back. If you eat just the right amount of food, then your average internal energy remains constant. Whatever you lose to heat transfer and doing work is replaced by food, so that, in the long run, $\Delta U = 0$. If you overeat repeatedly, then ΔU is always positive, and your body stores this extra internal energy as fat. The reverse is true if you eat too little. If ΔU is negative for a few days, then the body metabolizes its own fat to maintain body temperature and do work that takes energy from the body. This process is how dieting produces weight loss.

Life is not always this simple, as any dieter knows. The body stores fat or metabolizes it only if energy intake changes for a period of several days. Once you have been on a major diet, the next one is less successful because your body alters the way it responds to low energy intake. Your basal metabolic rate (BMR) is the rate at which food is converted into heat transfer and work done while the body is at complete rest. The body adjusts its basal metabolic rate to partially compensate for over-eating or under-eating. The body will decrease the metabolic rate rather than eliminate its own fat to replace lost food intake. You will chill more easily and feel less energetic as a result of the lower metabolic rate, and you will not lose weight as fast as before. Exercise helps to lose weight, because it produces both heat transfer from your body and work, and raises your metabolic rate even when you are at rest. Weight loss is also aided by the quite low efficiency of the body in converting internal energy to work, so that the loss of internal energy resulting from doing work is much greater than the work done. It should be noted, however, that living systems are not in thermalequilibrium.

The body provides us with an excellent indication that many thermodynamic processes are *irreversible*. An irreversible process can go in one direction but not the reverse, under a given set of conditions. For

example, although body fat can be converted to do work and produce heat transfer, work done on the body and heat transfer into it cannot be converted to body fat. Otherwise, we could skip lunch by sunning ourselves or by walking down stairs. Another example of an irreversible thermodynamic process is photosynthesis. This process is the intake of one form of energy—light—by plants and its conversion to chemical potential energy. Both applications of the first law of thermodynamics are illustrated in [\[link\]](#). One great advantage of conservation laws such as the first law of thermodynamics is that they accurately describe the beginning and ending points of complex processes, such as metabolism and photosynthesis, without regard to the complications in between. [\[link\]](#) presents a summary of terms relevant to the first law of thermodynamics.



(a) The first law of thermodynamics applied to metabolism. Heat transferred out of the body (Q) and work done by the body (W) remove internal energy, while food intake replaces it. (Food intake may be considered as work done on the body.) (b) Plants convert part of the radiant heat transfer in sunlight to stored chemical energy, a process called photosynthesis.

Term	Definition
U	Internal energy—the sum of the kinetic and potential energies of a system’s atoms and molecules. Can be divided into many subcategories, such as thermal and chemical energy. Depends only on the state of a system (such as its P , V , and T), not on how the energy entered the system. Change in internal energy is path independent.
Q	Heat—energy transferred because of a temperature difference. Characterized by random molecular motion. Highly dependent on path. Q entering a system is positive.
W	Work—energy transferred by a force moving through a distance. An organized, orderly process. Path dependent. W done by a system (either against an external force or to increase the volume of the system) is positive.

Summary of Terms for the First Law of Thermodynamics, $\Delta U = Q - W$

Section Summary

- The first law of thermodynamics is given as $\Delta U = Q - W$, where ΔU is the change in internal energy of a system, Q is the net heat transfer (the sum of all heat transfer into and out of the system), and W is the net work done (the sum of all work done on or by the system).
- Both Q and W are energy in transit; only ΔU represents an independent quantity capable of being stored.

- The internal energy U of a system depends only on the state of the system and not how it reached that state.
- Metabolism of living organisms, and photosynthesis of plants, are specialized types of heat transfer, doing work, and internal energy of systems.

Conceptual Questions

Exercise:

Problem:

Describe the photo of the tea kettle at the beginning of this section in terms of heat transfer, work done, and internal energy. How is heat being transferred? What is the work done and what is doing it? How does the kettle maintain its internal energy?

Exercise:

Problem:

The first law of thermodynamics and the conservation of energy, as discussed in [Conservation of Energy](#), are clearly related. How do they differ in the types of energy considered?

Exercise:

Problem:

Heat transfer Q and work done W are always energy in transit, whereas internal energy U is energy stored in a system. Give an example of each type of energy, and state specifically how it is either in transit or resides in a system.

Exercise:

Problem:

How do heat transfer and internal energy differ? In particular, which can be stored as such in a system and which cannot?

Exercise:

Problem:

If you run down some stairs and stop, what happens to your kinetic energy and your initial gravitational potential energy?

Exercise:**Problem:**

Give an explanation of how food energy (calories) can be viewed as molecular potential energy (consistent with the atomic and molecular definition of internal energy).

Exercise:**Problem:**

Identify the type of energy transferred to your body in each of the following as either internal energy, heat transfer, or doing work: (a) basking in sunlight; (b) eating food; (c) riding an elevator to a higher floor.

Problems & Exercises**Exercise:****Problem:**

What is the change in internal energy of a car if you put 12.0 gal of gasoline into its tank? The energy content of gasoline is $1.3 \times 10^8 \text{ J/gal}$. All other factors, such as the car's temperature, are constant.

Solution:

$$1.6 \times 10^9 \text{ J}$$

Exercise:

Problem:

How much heat transfer occurs from a system, if its internal energy decreased by 150 J while it was doing 30.0 J of work?

Exercise:**Problem:**

A system does 1.80×10^8 J of work while 7.50×10^8 J of heat transfer occurs to the environment. What is the change in internal energy of the system assuming no other changes (such as in temperature or by the addition of fuel)?

Solution:

$$-9.30 \times 10^8 \text{ J}$$

Exercise:**Problem:**

What is the change in internal energy of a system which does 4.50×10^5 J of work while 3.00×10^6 J of heat transfer occurs into the system, and 8.00×10^6 J of heat transfer occurs to the environment?

Exercise:**Problem:**

Suppose a woman does 500 J of work and 9500 J of heat transfer occurs into the environment in the process. (a) What is the decrease in her internal energy, assuming no change in temperature or consumption of food? (That is, there is no other energy transfer.) (b) What is her efficiency?

Solution:

(a) -1.0×10^4 J , or -2.39 kcal

(b) 5.00%

Exercise:**Problem:**

(a) How much food energy will a man metabolize in the process of doing 35.0 kJ of work with an efficiency of 5.00%? (b) How much heat transfer occurs to the environment to keep his temperature constant? Explicitly show how you follow the steps in the Problem-Solving Strategy for thermodynamics found in [Problem-Solving Strategies for Thermodynamics](#).

Exercise:**Problem:**

(a) What is the average metabolic rate in watts of a man who metabolizes 10,500 kJ of food energy in one day? (b) What is the maximum amount of work in joules he can do without breaking down fat, assuming a maximum efficiency of 20.0%? (c) Compare his work output with the daily output of a 187-W (0.250-horsepower) motor.

Solution:

(a) 122 W

(b) 2.10×10^6 J

(c) Work done by the motor is 1.61×10^7 J ;thus the motor produces 7.67 times the work done by the man

Exercise:**Problem:**

(a) How long will the energy in a 1470-kJ (350-kcal) cup of yogurt last in a woman doing work at the rate of 150 W with an efficiency of 20.0% (such as in leisurely climbing stairs)? (b) Does the time found in part (a) imply that it is easy to consume more food energy than you can reasonably expect to work off with exercise?

Exercise:

Problem:

(a) A woman climbing the Washington Monument metabolizes 6.00×10^2 kJ of food energy. If her efficiency is 18.0%, how much heat transfer occurs to the environment to keep her temperature constant? (b) Discuss the amount of heat transfer found in (a). Is it consistent with the fact that you quickly warm up when exercising?

Solution:

(a) 492 kJ

(b) This amount of heat is consistent with the fact that you warm quickly when exercising. Since the body is inefficient, the excess heat produced must be dissipated through sweating, breathing, etc.

Glossary

first law of thermodynamics

states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system

internal energy

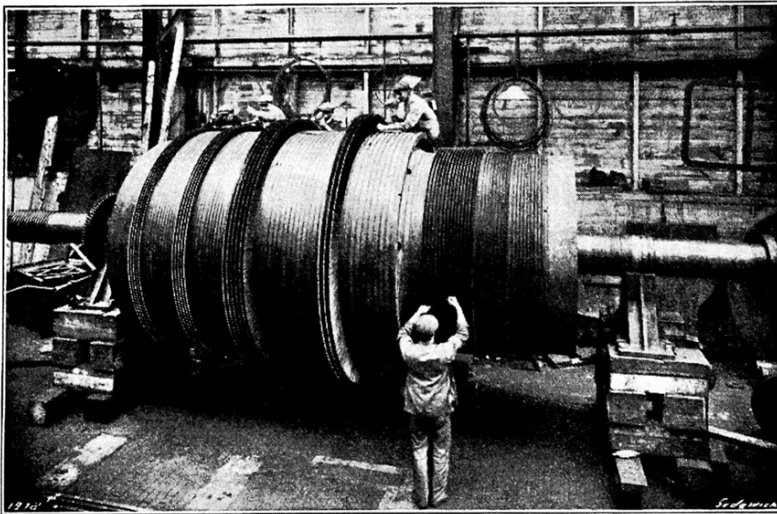
the sum of the kinetic and potential energies of a system's atoms and molecules

human metabolism

conversion of food into heat transfer, work, and stored fat

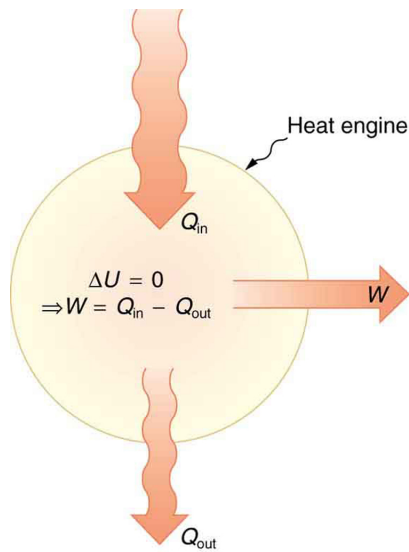
The First Law of Thermodynamics and Some Simple Processes

- Describe the processes of a simple heat engine.
- Explain the differences among the simple thermodynamic processes— isobaric, isochoric, isothermal, and adiabatic.
- Calculate total work done in a cyclical thermodynamic process.

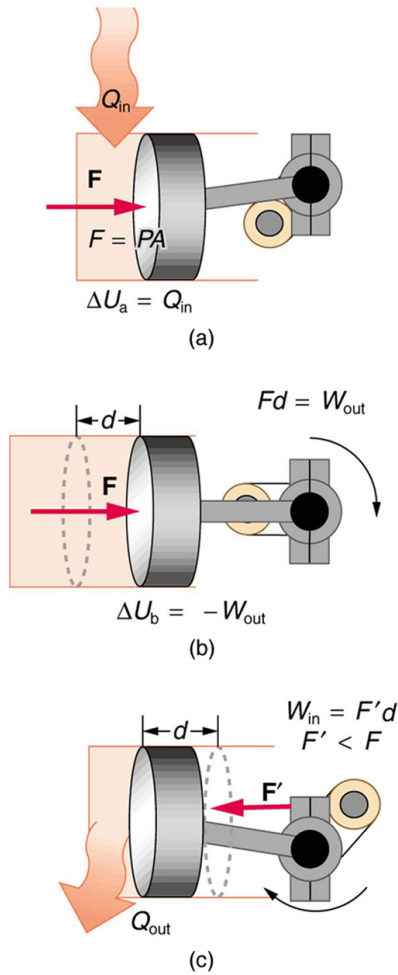


Beginning with the Industrial Revolution, humans have harnessed power through the use of the first law of thermodynamics, before we even understood it completely. This photo, of a steam engine at the Turbinia Works, dates from 1911, a mere 61 years after the first explicit statement of the first law of thermodynamics by Rudolph Clausius. (credit: public domain; author unknown)

One of the most important things we can do with heat transfer is to use it to do work for us. Such a device is called a **heat engine**. Car engines and steam turbines that generate electricity are examples of heat engines. [\[link\]](#) shows schematically how the first law of thermodynamics applies to the typical heat engine.



Schematic representation of a heat engine, governed, of course, by the first law of thermodynamics. It is impossible to devise a system where $Q_{out} = 0$, that is, in which no heat transfer occurs to the environment.



(a) Heat transfer to the gas in a cylinder increases the internal energy of the gas, creating higher pressure and temperature. (b) The force exerted on the movable cylinder does work as the gas expands. Gas pressure and temperature decrease when it expands, indicating

that the gas's
internal energy has
been decreased by
doing work. (c)
Heat transfer to the
environment
further reduces
pressure in the gas
so that the piston
can be more easily
returned to its
starting position.

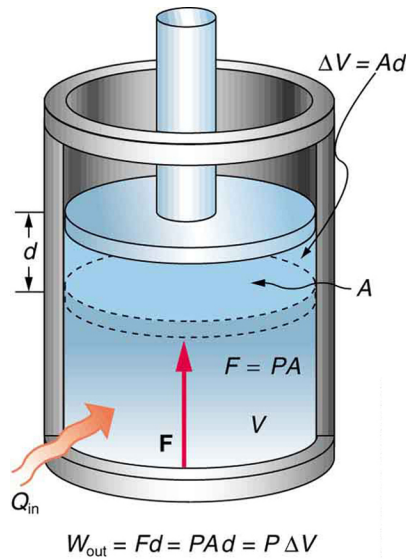
The illustrations above show one of the ways in which heat transfer does work. Fuel combustion produces heat transfer to a gas in a cylinder, increasing the pressure of the gas and thereby the force it exerts on a movable piston. The gas does work on the outside world, as this force moves the piston through some distance. Heat transfer to the gas cylinder results in work being done. To repeat this process, the piston needs to be returned to its starting point. Heat transfer now occurs from the gas to the surroundings so that its pressure decreases, and a force is exerted by the surroundings to push the piston back through some distance. Variations of this process are employed daily in hundreds of millions of heat engines. We will examine heat engines in detail in the next section. In this section, we consider some of the simpler underlying processes on which heat engines are based.

PV Diagrams and their Relationship to Work Done on or by a Gas

A process by which a gas does work on a piston at constant pressure is called an **isobaric process**. Since the pressure is constant, the force exerted is constant and the work done is given as

Equation:

$$P\Delta V.$$



An isobaric expansion of a gas requires heat transfer to keep the pressure constant. Since pressure is constant, the work done is $P\Delta V$.

Equation:

$$W = Fd$$

See the symbols as shown in [\[link\]](#). Now $F = PA$, and so

Equation:

$$W = PAd.$$

Because the volume of a cylinder is its cross-sectional area A times its length d , we see that $Ad = \Delta V$, the change in volume; thus,

Equation:

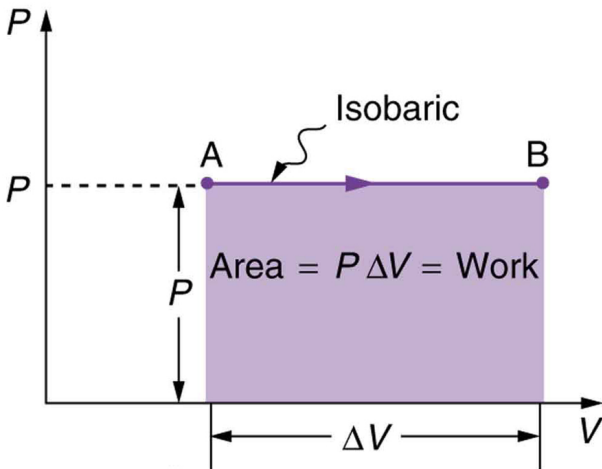
$$W = P\Delta V \text{ (isobaric process).}$$

Note that if ΔV is positive, then W is positive, meaning that work is done by the gas on the outside world.

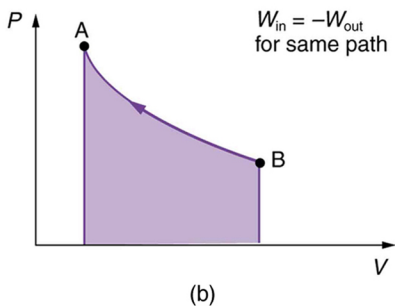
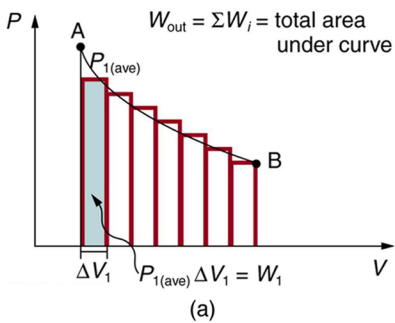
(Note that the pressure involved in this work that we've called P is the pressure of the gas *inside* the tank. If we call the pressure outside the tank P_{ext} , an expanding gas would be working *against* the external pressure; the work done would therefore be $W = -P_{\text{ext}}\Delta V$ (isobaric process). Many texts use this definition of work, and not the definition based on internal pressure, as the basis of the First Law of Thermodynamics. This definition reverses the sign conventions for work, and results in a statement of the first law that becomes $\Delta U = Q + W$.)

It is not surprising that $W = P\Delta V$, since we have already noted in our treatment of fluids that pressure is a type of potential energy per unit volume and that pressure in fact has units of energy divided by volume. We also noted in our discussion of the ideal gas law that PV has units of energy. In this case, some of the energy associated with pressure becomes work.

[\[link\]](#) shows a graph of pressure versus volume (that is, a PV diagram for an isobaric process. You can see in the figure that the work done is the area under the graph. This property of PV diagrams is very useful and broadly applicable: *the work done on or by a system in going from one state to another equals the area under the curve on a PV diagram.*



A graph of pressure versus volume for a constant-pressure, or isobaric, process, such as the one shown in [\[link\]](#). The area under the curve equals the work done by the gas, since $W = P\Delta V$.

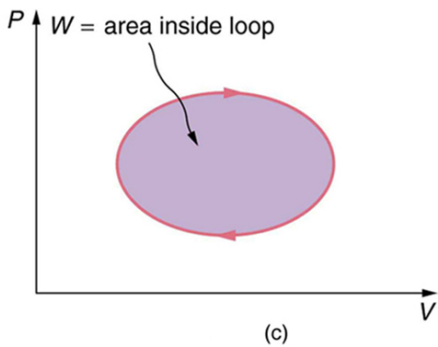
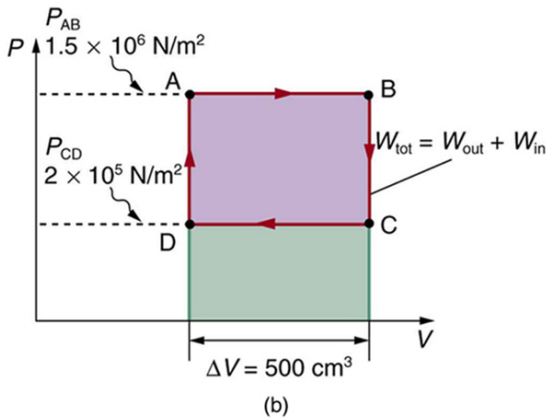
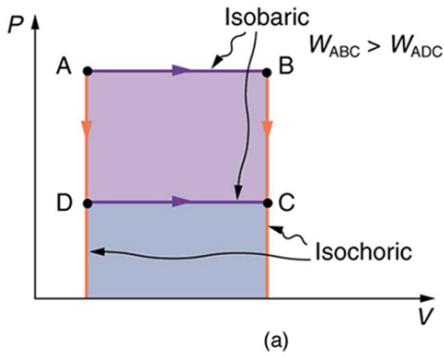


(a) A PV diagram in which pressure varies as well as volume. The work done for each interval is its average pressure times the change in volume, or the area under the curve over that interval. Thus the total area under the curve equals the total work done. (b) Work must be done on the system to follow the reverse path. This is interpreted as a negative area under the curve.

We can see where this leads by considering [\[link\]](#)(a), which shows a more general process in which both pressure and volume change. The area under the curve is closely approximated by dividing it into strips, each having an average constant pressure $P_{i(\text{ave})}$. The work done is $W_i = P_{i(\text{ave})}\Delta V_i$ for each strip, and the total work done is the sum of the W_i . Thus the total work done is the total area under the curve. If the path is reversed, as in [\[link\]](#)(b), then work is done on the system. The area under the curve in that case is negative, because ΔV is negative.

PV diagrams clearly illustrate that *the work done depends on the path taken and not just the endpoints*. This path dependence is seen in [\[link\]](#)(a), where

more work is done in going from A to C by the path via point B than by the path via point D. The vertical paths, where volume is constant, are called **isochoric** processes. Since volume is constant, $\Delta V = 0$, and no work is done in an isochoric process. Now, if the system follows the cyclical path ABCDA, as in [\[link\]](#)(b), then the total work done is the area inside the loop. The negative area below path CD subtracts, leaving only the area inside the rectangle. In fact, the work done in any cyclical process (one that returns to its starting point) is the area inside the loop it forms on a PV diagram, as [\[link\]](#)(c) illustrates for a general cyclical process. Note that the loop must be traversed in the clockwise direction for work to be positive—that is, for there to be a net work output.



(a) The work done in going from A to C depends on path. The work is greater for the path ABC than for the path ADC, because the former is at higher pressure. In both cases, the work done is the area under the path. This area is greater for path ABC. (b) The total work done in the cyclical process

ABCD is the area inside the loop, since the negative area below CD subtracts out, leaving just the area inside the rectangle. (The values given for the pressures and the change in volume are intended for use in the example below.) (c) The area inside any closed loop is the work done in the cyclical process. If the loop is traversed in a clockwise direction, W is positive—it is work done on the outside environment. If the loop is traveled in a counter-clockwise direction, W is negative—it is work that is done to the system.

Example:**Total Work Done in a Cyclical Process Equals the Area Inside the Closed Loop on a PV Diagram**

Calculate the total work done in the cyclical process ABCDA shown in [\[link\]](#)(b) by the following two methods to verify that work equals the area inside the closed loop on the PV diagram. (Take the data in the figure to be precise to three significant figures.) (a) Calculate the work done along each segment of the path and add these values to get the total work. (b) Calculate the area inside the rectangle ABCDA.

Strategy

To find the work along any path on a PV diagram, you use the fact that work is pressure times change in volume, or $W = P\Delta V$. So in part (a),

this value is calculated for each leg of the path around the closed loop.

Solution for (a)

The work along path AB is

Equation:

$$\begin{aligned}W_{AB} &= P_{AB}\Delta V_{AB} \\&= (1.50 \times 10^6 \text{ N/m}^2)(5.00 \times 10^{-4} \text{ m}^3) = 750 \text{ J}.\end{aligned}$$

Since the path BC is isochoric, $\Delta V_{BC} = 0$, and so $W_{BC} = 0$. The work along path CD is negative, since ΔV_{CD} is negative (the volume decreases). The work is

Equation:

$$\begin{aligned}W_{CD} &= P_{CD}\Delta V_{CD} \\&= (2.00 \times 10^5 \text{ N/m}^2)(-5.00 \times 10^{-4} \text{ m}^3) = -100 \text{ J}.\end{aligned}$$

Again, since the path DA is isochoric, $\Delta V_{DA} = 0$, and so $W_{DA} = 0$. Now the total work is

Equation:

$$\begin{aligned}W &= W_{AB} + W_{BC} + W_{CD} + W_{DA} \\&= 750 \text{ J} + 0 + (-100 \text{ J}) + 0 = 650 \text{ J}.\end{aligned}$$

Solution for (b)

The area inside the rectangle is its height times its width, or

Equation:

$$\begin{aligned}\text{area} &= (P_{AB} - P_{CD})\Delta V \\&= \left[(1.50 \times 10^6 \text{ N/m}^2) - (2.00 \times 10^5 \text{ N/m}^2) \right] (5.00 \times 10^{-4} \text{ m}^3) \\&= 650 \text{ J}.\end{aligned}$$

Thus,

Equation:

$$\text{area} = 650 \text{ J} = W.$$

Discussion

The result, as anticipated, is that the area inside the closed loop equals the work done. The area is often easier to calculate than is the work done along each path. It is also convenient to visualize the area inside different curves on PV diagrams in order to see which processes might produce the most work. Recall that work can be done to the system, or by the system, depending on the sign of W . A positive W is work that is done by the system on the outside environment; a negative W represents work done by the environment on the system.

[\[link\]](#)(a) shows two other important processes on a PV diagram. For comparison, both are shown starting from the same point A. The upper curve ending at point B is an **isothermal** process—that is, one in which temperature is kept constant. If the gas behaves like an ideal gas, as is often the case, and if no phase change occurs, then $PV = nRT$. Since T is constant, PV is a constant for an isothermal process. We ordinarily expect the temperature of a gas to decrease as it expands, and so we correctly suspect that heat transfer must occur from the surroundings to the gas to keep the temperature constant during an isothermal expansion. To show this more rigorously for the special case of a monatomic ideal gas, we note that the average kinetic energy of an atom in such a gas is given by

Equation:

$$\frac{1}{2}mv^2 = \frac{3}{2}kT.$$

The kinetic energy of the atoms in a monatomic ideal gas is its only form of internal energy, and so its total internal energy U is

Equation:

$$U = N\frac{1}{2}mv^2 = \frac{3}{2}NkT, \text{ (monatomic ideal gas),}$$

where N is the number of atoms in the gas. This relationship means that the internal energy of an ideal monatomic gas is constant during an isothermal process—that is, $\Delta U = 0$. If the internal energy does not change, then the net heat transfer into the gas must equal the net work done by the gas. That is, because $\Delta U = Q - W = 0$ here, $Q = W$. We must have just enough heat transfer to replace the work done. An isothermal

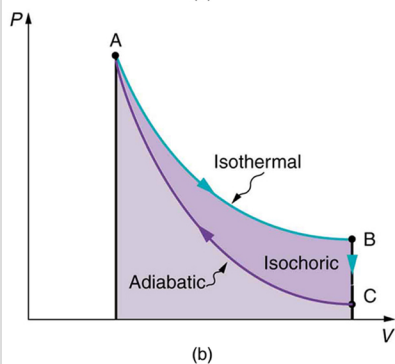
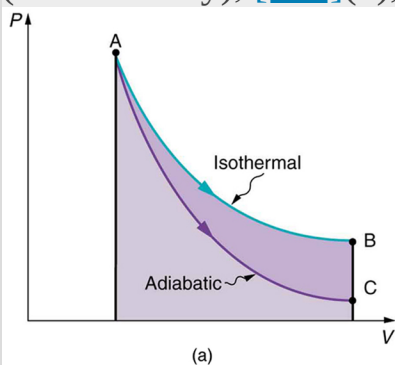
process is inherently slow, because heat transfer occurs continuously to keep the gas temperature constant at all times and must be allowed to spread through the gas so that there are no hot or cold regions.

Also shown in [\[link\]](#)(a) is a curve AC for an **adiabatic** process, defined to be one in which there is no heat transfer—that is, $Q = 0$. Processes that are nearly adiabatic can be achieved either by using very effective insulation or by performing the process so fast that there is little time for heat transfer. Temperature must decrease during an adiabatic expansion process, since work is done at the expense of internal energy:

Equation:

$$U = \frac{3}{2}NkT.$$

(You might have noted that a gas released into atmospheric pressure from a pressurized cylinder is substantially colder than the gas in the cylinder.) In fact, because $Q = 0$, $\Delta U = -W$ for an adiabatic process. Lower temperature results in lower pressure along the way, so that curve AC is lower than curve AB, and less work is done. If the path ABCA could be followed by cooling the gas from B to C at constant volume (isochorically), [\[link\]](#)(b), there would be a net work output.



(a) The upper curve is an isothermal process ($\Delta T = 0$), whereas the lower curve is an adiabatic process ($Q = 0$). Both start from the same point A, but the isothermal process does more work than the adiabatic because heat transfer into the gas takes place to keep its temperature constant. This keeps the pressure higher all along the isothermal path than along the adiabatic path, producing more work. The adiabatic path thus ends up with a lower pressure and temperature at point C, even though the final volume is the same as for the isothermal process.

(b) The cycle ABCA produces a net work output.

Reversible Processes

Both isothermal and adiabatic processes such as shown in [\[link\]](#) are reversible in principle. A **reversible process** is one in which both the system and its environment can return to exactly the states they were in by following the reverse path. The reverse isothermal and adiabatic paths are BA and CA, respectively. Real macroscopic processes are never exactly reversible. In the previous examples, our system is a gas (like that in [\[link\]](#)), and its environment is the piston, cylinder, and the rest of the universe. If there are any energy-dissipating mechanisms, such as friction or turbulence, then heat transfer to the environment occurs for either direction of the piston. So, for example, if the path BA is followed and there is friction, then the gas will be returned to its original state but the environment will not—it will have been heated in both directions. Reversibility requires the direction of heat transfer to reverse for the reverse path. Since dissipative mechanisms cannot be completely eliminated, real processes cannot be reversible.

There must be reasons that real macroscopic processes cannot be reversible. We can imagine them going in reverse. For example, heat transfer occurs spontaneously from hot to cold and never spontaneously the reverse. Yet it would not violate the first law of thermodynamics for this to happen. In fact, all spontaneous processes, such as bubbles bursting, never go in reverse. There is a second thermodynamic law that forbids them from going in reverse. When we study this law, we will learn something about nature and also find that such a law limits the efficiency of heat engines. We will find that heat engines with the greatest possible theoretical efficiency would have to use reversible processes, and even they cannot convert all heat transfer into doing work. [\[link\]](#) summarizes the simpler thermodynamic processes and their definitions.

Isobaric	Constant pressure $W = P\Delta V$
Isochoric	Constant volume $W = 0$
Isothermal	Constant temperature $Q = W$
Adiabatic	No heat transfer $Q = 0$

Summary of Simple Thermodynamic Processes

Note:

PhET Explorations: States of Matter

Watch different types of molecules form a solid, liquid, or gas. Add or remove heat and watch the phase change. Change the temperature or volume of a container and see a pressure-temperature diagram respond in real time. Relate the interaction potential to the forces between molecules.

https://phet.colorado.edu/sims/html/states-of-matter/latest/states-of-matter_en.html

Section Summary

- One of the important implications of the first law of thermodynamics is that machines can be harnessed to do work that humans previously

did by hand or by external energy supplies such as running water or the heat of the Sun. A machine that uses heat transfer to do work is known as a heat engine.

- There are several simple processes, used by heat engines, that flow from the first law of thermodynamics. Among them are the isobaric, isochoric, isothermal and adiabatic processes.
- These processes differ from one another based on how they affect pressure, volume, temperature, and heat transfer.
- If the work done is performed on the outside environment, work (W) will be a positive value. If the work done is done to the heat engine system, work (W) will be a negative value.
- Some thermodynamic processes, including isothermal and adiabatic processes, are reversible in theory; that is, both the thermodynamic system and the environment can be returned to their initial states. However, because of loss of energy owing to the second law of thermodynamics, complete reversibility does not work in practice.

Conceptual Questions

Exercise:

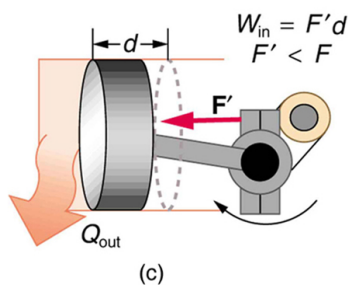
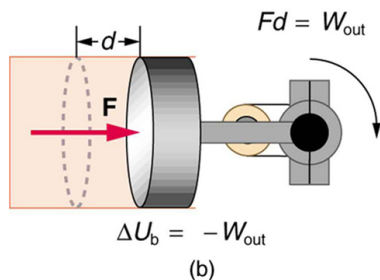
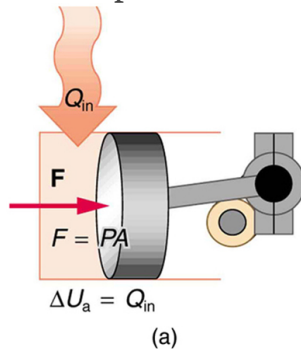
Problem:

A great deal of effort, time, and money has been spent in the quest for the so-called perpetual-motion machine, which is defined as a hypothetical machine that operates or produces useful work indefinitely and/or a hypothetical machine that produces more work or energy than it consumes. Explain, in terms of heat engines and the first law of thermodynamics, why or why not such a machine is likely to be constructed.

Exercise:

Problem:

One method of converting heat transfer into doing work is for heat transfer into a gas to take place, which expands, doing work on a piston, as shown in the figure below. (a) Is the heat transfer converted directly to work in an isobaric process, or does it go through another form first? Explain your answer. (b) What about in an isothermal process? (c) What about in an adiabatic process (where heat transfer occurred prior to the adiabatic process)?



Exercise:

Problem:

Would the previous question make any sense for an isochoric process? Explain your answer.

Exercise:**Problem:**

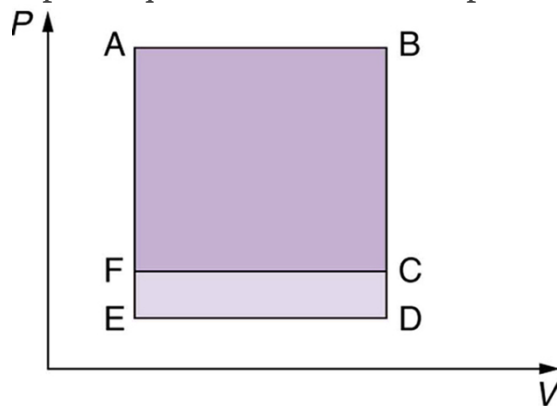
We ordinarily say that $\Delta U = 0$ for an isothermal process. Does this assume no phase change takes place? Explain your answer.

Exercise:**Problem:**

The temperature of a rapidly expanding gas decreases. Explain why in terms of the first law of thermodynamics. (Hint: Consider whether the gas does work and whether heat transfer occurs rapidly into the gas through conduction.)

Exercise:**Problem:**

Which cyclical process represented by the two closed loops, ABCFA and ABDEA, on the PV diagram in the figure below produces the greatest *net* work? Is that process also the one with the smallest work input required to return it to point A? Explain your responses.



The two cyclical processes shown on this PV diagram start with and return the system to the conditions at point A, but they follow

different paths and produce different amounts of work.

Exercise:

Problem:

A real process may be nearly adiabatic if it occurs over a very short time. How does the short time span help the process to be adiabatic?

Exercise:

Problem:

It is unlikely that a process can be isothermal unless it is a very slow process. Explain why. Is the same true for isobaric and isochoric processes? Explain your answer.

Problem Exercises

Exercise:

Problem:

A car tire contains 0.0380 m^3 of air at a pressure of $2.20 \times 10^5 \text{ N/m}^2$ (about 32 psi). How much more internal energy does this gas have than the same volume has at zero gauge pressure (which is equivalent to normal atmospheric pressure)?

Solution:

$$6.77 \times 10^3 \text{ J}$$

Exercise:

Problem:

A helium-filled toy balloon has a gauge pressure of 0.200 atm and a volume of 10.0 L. How much greater is the internal energy of the helium in the balloon than it would be at zero gauge pressure?

Exercise:**Problem:**

Steam to drive an old-fashioned steam locomotive is supplied at a constant gauge pressure of $1.75 \times 10^6 \text{ N/m}^2$ (about 250 psi) to a piston with a 0.200-m radius. (a) By calculating $P\Delta V$, find the work done by the steam when the piston moves 0.800 m. Note that this is the net work output, since gauge pressure is used. (b) Now find the amount of work by calculating the force exerted times the distance traveled. Is the answer the same as in part (a)?

Solution:

(a) $W = P\Delta V = 1.76 \times 10^5 \text{ J}$

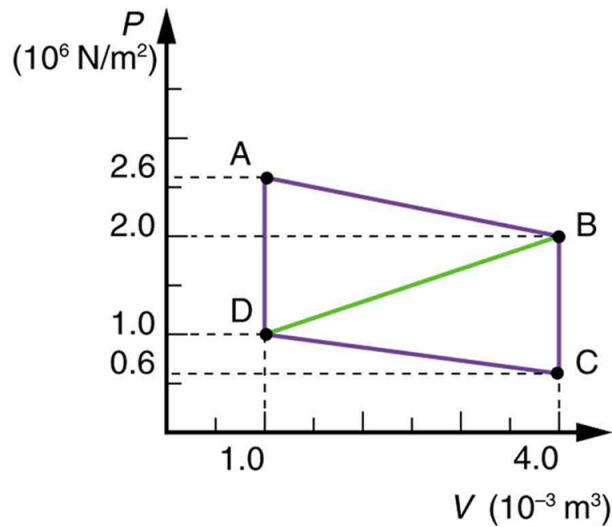
(b) $W = Fd = 1.76 \times 10^5 \text{ J}$. Yes, the answer is the same.

Exercise:**Problem:**

A hand-driven tire pump has a piston with a 2.50-cm diameter and a maximum stroke of 30.0 cm. (a) How much work do you do in one stroke if the average gauge pressure is $2.40 \times 10^5 \text{ N/m}^2$ (about 35 psi)? (b) What average force do you exert on the piston, neglecting friction and gravitational force?

Exercise:**Problem:**

Calculate the net work output of a heat engine following path ABCDA in the figure below.



Solution:

$$W = 4.5 \times 10^3 \text{ J}$$

Exercise:

Problem:

What is the net work output of a heat engine that follows path ABDA in the figure above, with a straight line from B to D? Why is the work output less than for path ABCDA? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Thermodynamics](#).

Exercise:

Problem: Unreasonable Results

What is wrong with the claim that a cyclical heat engine does 4.00 kJ of work on an input of 24.0 kJ of heat transfer while 16.0 kJ of heat transfers to the environment?

Solution:

W is not equal to the difference between the heat input and the heat output.

Exercise:

Problem:

- (a) A cyclical heat engine, operating between temperatures of 450°C and 150°C produces 4.00 MJ of work on a heat transfer of 5.00 MJ into the engine. How much heat transfer occurs to the environment?
(b) What is unreasonable about the engine? (c) Which premise is unreasonable?

Exercise:**Problem: Construct Your Own Problem**

Consider a car's gasoline engine. Construct a problem in which you calculate the maximum efficiency this engine can have. Among the things to consider are the effective hot and cold reservoir temperatures. Compare your calculated efficiency with the actual efficiency of car engines.

Exercise:**Problem: Construct Your Own Problem**

Consider a car trip into the mountains. Construct a problem in which you calculate the overall efficiency of the car for the trip as a ratio of kinetic and potential energy gained to fuel consumed. Compare this efficiency to the thermodynamic efficiency quoted for gasoline engines and discuss why the thermodynamic efficiency is so much greater. Among the factors to be considered are the gain in altitude and speed, the mass of the car, the distance traveled, and typical fuel economy.

Glossary

heat engine

a machine that uses heat transfer to do work

isobaric process

constant-pressure process in which a gas does work

isochoric process

a constant-volume process

isothermal process

a constant-temperature process

adiabatic process

a process in which no heat transfer takes place

reversible process

a process in which both the heat engine system and the external environment theoretically can be returned to their original states

Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency

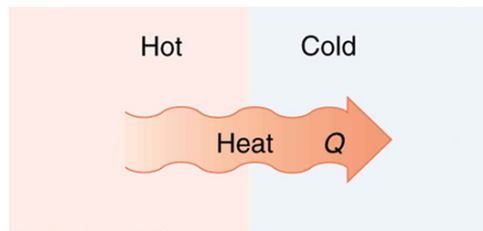
- State the expressions of the second law of thermodynamics.
- Calculate the efficiency and carbon dioxide emission of a coal-fired electricity plant, using second law characteristics.
- Describe and define the Otto cycle.



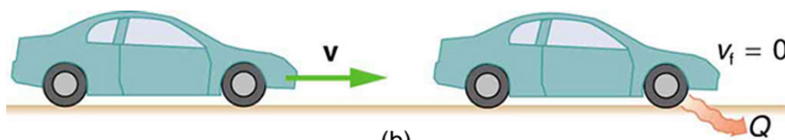
These ice floes melt during the Arctic summer. Some of them refreeze in the winter, but the second law of thermodynamics predicts that it would be extremely unlikely for the water molecules contained in these particular floes to reform the distinctive alligator-like shape they formed when the picture was taken in the summer of 2009. (credit: Patrick Kelley, U.S. Coast Guard, U.S. Geological Survey)

The second law of thermodynamics deals with the direction taken by spontaneous processes. Many processes occur spontaneously in one direction only—that is, they are irreversible, under a given set of

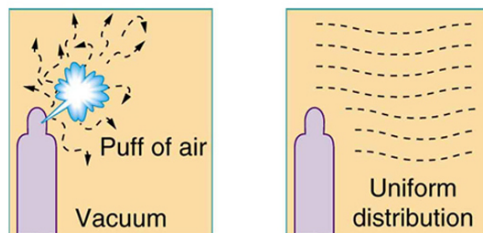
conditions. Although irreversibility is seen in day-to-day life—a broken glass does not resume its original state, for instance—complete irreversibility is a statistical statement that cannot be seen during the lifetime of the universe. More precisely, an **irreversible process** is one that depends on path. If the process can go in only one direction, then the reverse path differs fundamentally and the process cannot be reversible. For example, as noted in the previous section, heat involves the transfer of energy from higher to lower temperature. A cold object in contact with a hot one never gets colder, transferring heat to the hot object and making it hotter. Furthermore, mechanical energy, such as kinetic energy, can be completely converted to thermal energy by friction, but the reverse is impossible. A hot stationary object never spontaneously cools off and starts moving. Yet another example is the expansion of a puff of gas introduced into one corner of a vacuum chamber. The gas expands to fill the chamber, but it never regroups in the corner. The random motion of the gas molecules could take them all back to the corner, but this is never observed to happen. (See [\[link\]](#).)



(a)



(b)



(c)

Examples of one-way processes in nature.

(a) Heat transfer occurs spontaneously from hot to cold and not from cold to hot. (b) The brakes of this car convert its kinetic energy to heat transfer to the environment. The reverse process is impossible. (c) The burst of gas let into this vacuum chamber quickly expands to uniformly fill every part of the chamber. The random motions of the gas molecules will never return them to the corner.

The fact that certain processes never occur suggests that there is a law forbidding them to occur. The first law of thermodynamics would allow them to occur—none of those processes violate conservation of energy. The law that forbids these processes is called the second law of thermodynamics. We shall see that the second law can be stated in many ways that may seem different, but which in fact are equivalent. Like all natural laws, the second law of thermodynamics gives insights into nature, and its several statements imply that it is broadly applicable, fundamentally affecting many apparently disparate processes.

The already familiar direction of heat transfer from hot to cold is the basis of our first version of the **second law of thermodynamics**.

Note:

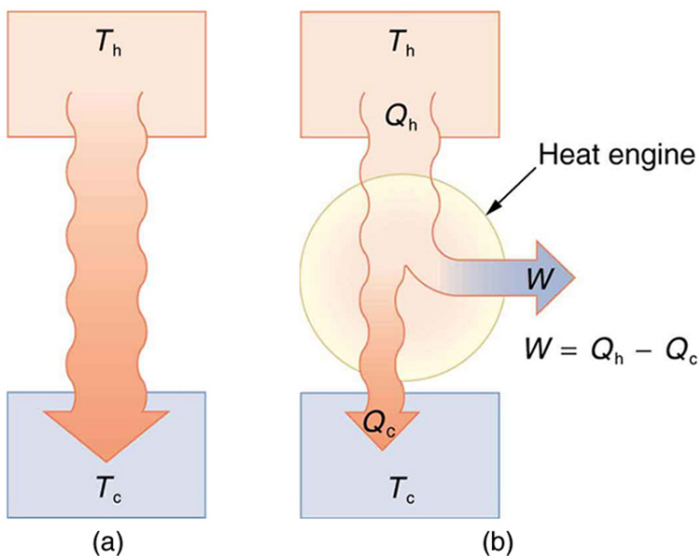
The Second Law of Thermodynamics (first expression)

Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction.

Another way of stating this: It is impossible for any process to have as its sole result heat transfer from a cooler to a hotter object.

Heat Engines

Now let us consider a device that uses heat transfer to do work. As noted in the previous section, such a device is called a heat engine, and one is shown schematically in [\[link\]](#)(b). Gasoline and diesel engines, jet engines, and steam turbines are all heat engines that do work by using part of the heat transfer from some source. Heat transfer from the hot object (or hot reservoir) is denoted as Q_h , while heat transfer into the cold object (or cold reservoir) is Q_c , and the work done by the engine is W . The temperatures of the hot and cold reservoirs are T_h and T_c , respectively.



- (a) Heat transfer occurs spontaneously from a hot object to a cold one, consistent with the second law of thermodynamics. (b) A heat engine, represented here by a circle, uses part of the heat transfer to do work. The hot and cold objects are called the hot and cold reservoirs. Q_h is the heat transfer out of the hot reservoir, W is the work output, and Q_c is the heat transfer into the cold reservoir.

Because the hot reservoir is heated externally, which is energy intensive, it is important that the work is done as efficiently as possible. In fact, we would like W to equal Q_h , and for there to be no heat transfer to the environment ($Q_c = 0$). Unfortunately, this is impossible. The **second law of thermodynamics** also states, with regard to using heat transfer to do work (the second expression of the second law):

Note:

The Second Law of Thermodynamics (second expression)

It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.

A **cyclical process** brings a system, such as the gas in a cylinder, back to its original state at the end of every cycle. Most heat engines, such as reciprocating piston engines and rotating turbines, use cyclical processes. The second law, just stated in its second form, clearly states that such engines cannot have perfect conversion of heat transfer into work done. Before going into the underlying reasons for the limits on converting heat transfer into work, we need to explore the relationships among W , Q_h , and Q_c , and to define the efficiency of a cyclical heat engine. As noted, a cyclical process brings the system back to its original condition at the end of every cycle. Such a system's internal energy U is the same at the beginning and end of every cycle—that is, $\Delta U = 0$. The first law of thermodynamics states that

Equation:

$$\Delta U = Q - W,$$

where Q is the *net* heat transfer during the cycle ($Q = Q_h - Q_c$) and W is the net work done by the system. Since $\Delta U = 0$ for a complete cycle, we

have

Equation:

$$0 = Q - W,$$

so that

Equation:

$$W = Q.$$

Thus the net work done by the system equals the net heat transfer into the system, or

Equation:

$$W = Q_h - Q_c \text{ (cyclical process),}$$

just as shown schematically in [\[link\]](#)(b). The problem is that in all processes, there is some heat transfer Q_c to the environment—and usually a very significant amount at that.

In the conversion of energy to work, we are always faced with the problem of getting less out than we put in. We define *conversion efficiency* Eff to be the ratio of useful work output to the energy input (or, in other words, the ratio of what we get to what we spend). In that spirit, we define the efficiency of a heat engine to be its net work output W divided by heat transfer to the engine Q_h ; that is,

Equation:

$$\text{Eff} = \frac{W}{Q_h}.$$

Since $W = Q_h - Q_c$ in a cyclical process, we can also express this as

Equation:

$$\text{Eff} = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h} \text{ (cyclical process),}$$

making it clear that an efficiency of 1, or 100%, is possible only if there is no heat transfer to the environment ($Q_c = 0$). Note that all Q s are positive. The direction of heat transfer is indicated by a plus or minus sign. For example, Q_c is out of the system and so is preceded by a minus sign.

Example:

Daily Work Done by a Coal-Fired Power Station, Its Efficiency and Carbon Dioxide Emissions

A coal-fired power station is a huge heat engine. It uses heat transfer from burning coal to do work to turn turbines, which are used to generate electricity. In a single day, a large coal power station has 2.50×10^{14} J of heat transfer from coal and 1.48×10^{14} J of heat transfer into the environment. (a) What is the work done by the power station? (b) What is the efficiency of the power station? (c) In the combustion process, the following chemical reaction occurs: $C + O_2 \rightarrow CO_2$. This implies that every 12 kg of coal puts $12 \text{ kg} + 16 \text{ kg} + 16 \text{ kg} = 44 \text{ kg}$ of carbon dioxide into the atmosphere. Assuming that 1 kg of coal can provide 2.5×10^6 J of heat transfer upon combustion, how much CO_2 is emitted per day by this power plant?

Strategy for (a)

We can use $W = Q_h - Q_c$ to find the work output W , assuming a cyclical process is used in the power station. In this process, water is boiled under pressure to form high-temperature steam, which is used to run steam turbine-generators, and then condensed back to water to start the cycle again.

Solution for (a)

Work output is given by:

Equation:

$$W = Q_h - Q_c.$$

Substituting the given values:

Equation:

$$\begin{aligned} W &= 2.50 \times 10^{14} \text{ J} - 1.48 \times 10^{14} \text{ J} \\ &= 1.02 \times 10^{14} \text{ J}. \end{aligned}$$

Strategy for (b)

The efficiency can be calculated with $\text{Eff} = \frac{W}{Q_h}$ since Q_h is given and work W was found in the first part of this example.

Solution for (b)

Efficiency is given by: $\text{Eff} = \frac{W}{Q_h}$. The work W was just found to be $1.02 \times 10^{14} \text{ J}$, and Q_h is given, so the efficiency is

Equation:

$$\begin{aligned} \text{Eff} &= \frac{1.02 \times 10^{14} \text{ J}}{2.50 \times 10^{14} \text{ J}} \\ &= 0.408, \text{ or } 40.8\% \end{aligned}$$

Strategy for (c)

The daily consumption of coal is calculated using the information that each day there is $2.50 \times 10^{14} \text{ J}$ of heat transfer from coal. In the combustion process, we have $\text{C} + \text{O}_2 \rightarrow \text{CO}_2$. So every 12 kg of coal puts 12 kg + 16 kg + 16 kg = 44 kg of CO_2 into the atmosphere.

Solution for (c)

The daily coal consumption is

Equation:

$$\frac{2.50 \times 10^{14} \text{ J}}{2.50 \times 10^6 \text{ J/kg}} = 1.0 \times 10^8 \text{ kg}.$$

Assuming that the coal is pure and that all the coal goes toward producing carbon dioxide, the carbon dioxide produced per day is

Equation:

$$1.0 \times 10^8 \text{ kg coal} \times \frac{44 \text{ kg CO}_2}{12 \text{ kg coal}} = 3.7 \times 10^8 \text{ kg CO}_2.$$

This is 370,000 metric tons of CO_2 produced every day.

Discussion

If all the work output is converted to electricity in a period of one day, the average power output is 1180 MW (this is left to you as an end-of-chapter problem). This value is about the size of a large-scale conventional power plant. The efficiency found is acceptably close to the value of 42% given for coal power stations. It means that fully 59.2% of the energy is heat transfer to the environment, which usually results in warming lakes, rivers, or the ocean near the power station, and is implicated in a warming planet generally. While the laws of thermodynamics limit the efficiency of such plants—including plants fired by nuclear fuel, oil, and natural gas—the heat transfer to the environment could be, and sometimes is, used for heating homes or for industrial processes. The generally low cost of energy has not made it economical to make better use of the waste heat transfer from most heat engines. Coal-fired power plants produce the greatest amount of CO_2 per unit energy output (compared to natural gas or oil), making coal the least efficient fossil fuel.

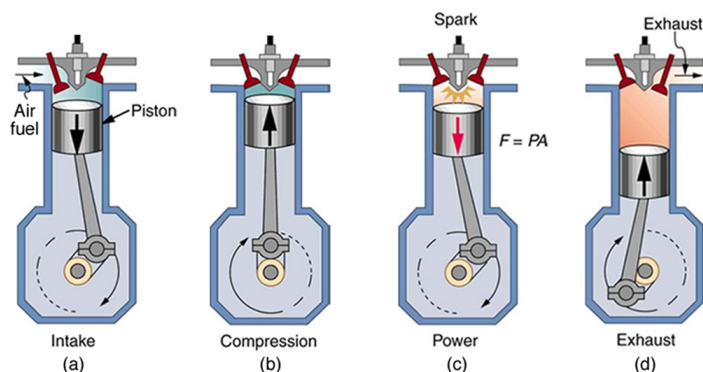
With the information given in [\[link\]](#), we can find characteristics such as the efficiency of a heat engine without any knowledge of how the heat engine operates, but looking further into the mechanism of the engine will give us greater insight. [\[link\]](#) illustrates the operation of the common four-stroke gasoline engine. The four steps shown complete this heat engine's cycle, bringing the gasoline-air mixture back to its original condition.

The **Otto cycle** shown in [\[link\]](#)(a) is used in four-stroke internal combustion engines, although in fact the true Otto cycle paths do not correspond exactly to the strokes of the engine.

The adiabatic process AB corresponds to the nearly adiabatic compression stroke of the gasoline engine. In both cases, work is done on the system (the gas mixture in the cylinder), increasing its temperature and pressure. Along path BC of the Otto cycle, heat transfer Q_h into the gas occurs at constant volume, causing a further increase in pressure and temperature. This process corresponds to burning fuel in an internal combustion engine, and takes place so rapidly that the volume is nearly constant. Path CD in the Otto cycle is an adiabatic expansion that does work on the outside world,

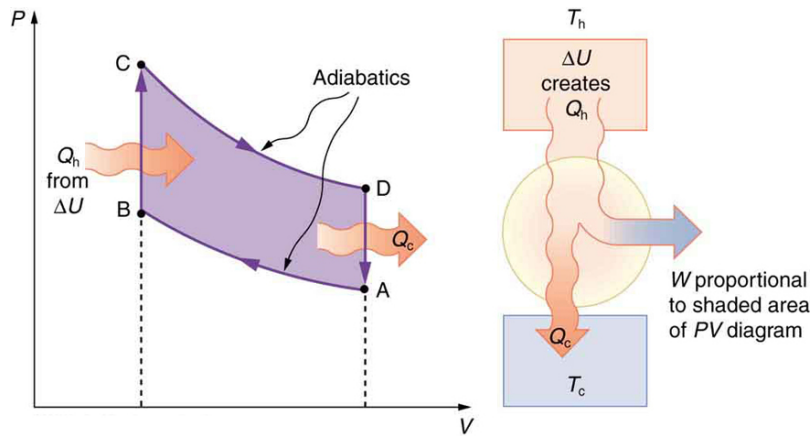
just as the power stroke of an internal combustion engine does in its nearly adiabatic expansion. The work done by the system along path CD is greater than the work done on the system along path AB, because the pressure is greater, and so there is a net work output. Along path DA in the Otto cycle, heat transfer Q_c from the gas at constant volume reduces its temperature and pressure, returning it to its original state. In an internal combustion engine, this process corresponds to the exhaust of hot gases and the intake of an air-gasoline mixture at a considerably lower temperature. In both cases, heat transfer into the environment occurs along this final path.

The net work done by a cyclical process is the area inside the closed path on a PV diagram, such as that inside path ABCDA in [\[link\]](#). Note that in every imaginable cyclical process, it is absolutely necessary for heat transfer from the system to occur in order to get a net work output. In the Otto cycle, heat transfer occurs along path DA. If no heat transfer occurs, then the return path is the same, and the net work output is zero. The lower the temperature on the path AB, the less work has to be done to compress the gas. The area inside the closed path is then greater, and so the engine does more work and is thus more efficient. Similarly, the higher the temperature along path CD, the more work output there is. (See [\[link\]](#).) So efficiency is related to the temperatures of the hot and cold reservoirs. In the next section, we shall see what the absolute limit to the efficiency of a heat engine is, and how it is related to temperature.



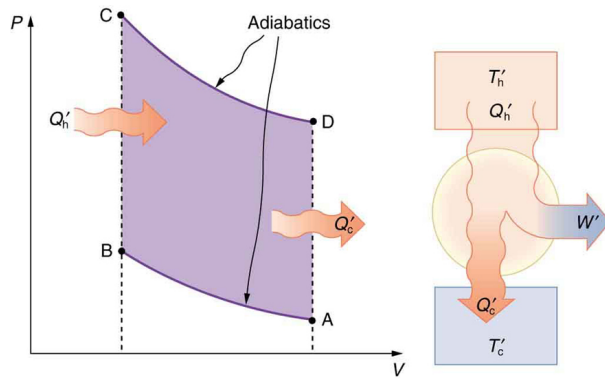
In the four-stroke internal combustion gasoline engine, heat transfer into

work takes place in the cyclical process shown here. The piston is connected to a rotating crankshaft, which both takes work out of and does work on the gas in the cylinder. (a) Air is mixed with fuel during the intake stroke. (b) During the compression stroke, the air-fuel mixture is rapidly compressed in a nearly adiabatic process, as the piston rises with the valves closed. Work is done on the gas. (c) The power stroke has two distinct parts. First, the air-fuel mixture is ignited, converting chemical potential energy into thermal energy almost instantaneously, which leads to a great increase in pressure. Then the piston descends, and the gas does work by exerting a force through a distance in a nearly adiabatic process. (d) The exhaust stroke expels the hot gas to prepare the engine for another cycle, starting again with the intake stroke.



PV diagram for a simplified Otto cycle, analogous to that employed in an internal combustion engine. Point A corresponds to the start of the compression stroke of an internal combustion engine. Paths AB and CD are adiabatic and correspond to the compression and power strokes of an internal combustion engine, respectively.

Paths BC and DA are isochoric and accomplish similar results to the ignition and exhaust-intake portions, respectively, of the internal combustion engine's cycle. Work is done on the gas along path AB, but more work is done by the gas along path CD, so that there is a net work output.



This Otto cycle produces a greater work output than the one in [\[link\]](#), because the starting temperature of path CD is higher and the starting temperature of path AB is lower. The area inside the loop is greater, corresponding to greater net work output.

Section Summary

- The two expressions of the second law of thermodynamics are: (i) Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction; and (ii) It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.
- Irreversible processes depend on path and do not return to their original state. Cyclical processes are processes that return to their original state at the end of every cycle.
- In a cyclical process, such as a heat engine, the net work done by the system equals the net heat transfer into the system, or $W = Q_h - Q_c$, where Q_h is the heat transfer from the hot object (hot reservoir), and Q_c is the heat transfer into the cold object (cold reservoir).

- Efficiency can be expressed as $\text{Eff} = \frac{W}{Q_h}$, the ratio of work output divided by the amount of energy input.
- The four-stroke gasoline engine is often explained in terms of the Otto cycle, which is a repeating sequence of processes that convert heat into work.

Conceptual Questions

Exercise:

Problem:

Imagine you are driving a car up Pike's Peak in Colorado. To raise a car weighing 1000 kilograms a distance of 100 meters would require about a million joules. You could raise a car 12.5 kilometers with the energy in a gallon of gas. Driving up Pike's Peak (a mere 3000-meter climb) should consume a little less than a quart of gas. But other considerations have to be taken into account. Explain, in terms of efficiency, what factors may keep you from realizing your ideal energy use on this trip.

Exercise:

Problem:

Is a temperature difference necessary to operate a heat engine? State why or why not.

Exercise:

Problem:

Definitions of efficiency vary depending on how energy is being converted. Compare the definitions of efficiency for the human body and heat engines. How does the definition of efficiency in each relate to the type of energy being converted into doing work?

Exercise:

Problem:

Why—other than the fact that the second law of thermodynamics says reversible engines are the most efficient—should heat engines employing reversible processes be more efficient than those employing irreversible processes? Consider that dissipative mechanisms are one cause of irreversibility.

Problem Exercises**Exercise:****Problem:**

A certain heat engine does 10.0 kJ of work and 8.50 kJ of heat transfer occurs to the environment in a cyclical process. (a) What was the heat transfer into this engine? (b) What was the engine's efficiency?

Solution:

(a) 18.5 kJ

(b) 54.1%

Exercise:**Problem:**

With 2.56×10^6 J of heat transfer into this engine, a given cyclical heat engine can do only 1.50×10^5 J of work. (a) What is the engine's efficiency? (b) How much heat transfer to the environment takes place?

Exercise:

Problem:

(a) What is the work output of a cyclical heat engine having a 22.0% efficiency and 6.00×10^9 J of heat transfer into the engine? (b) How much heat transfer occurs to the environment?

Solution:

(a) 1.32×10^9 J

(b) 4.68×10^9 J

Exercise:**Problem:**

(a) What is the efficiency of a cyclical heat engine in which 75.0 kJ of heat transfer occurs to the environment for every 95.0 kJ of heat transfer into the engine? (b) How much work does it produce for 100 kJ of heat transfer into the engine?

Exercise:**Problem:**

The engine of a large ship does 2.00×10^8 J of work with an efficiency of 5.00%. (a) How much heat transfer occurs to the environment? (b) How many barrels of fuel are consumed, if each barrel produces 6.00×10^9 J of heat transfer when burned?

Solution:

(a) 3.80×10^9 J

(b) 0.667 barrels

Exercise:

Problem:

(a) How much heat transfer occurs to the environment by an electrical power station that uses 1.25×10^{14} J of heat transfer into the engine with an efficiency of 42.0%? (b) What is the ratio of heat transfer to the environment to work output? (c) How much work is done?

Exercise:**Problem:**

Assume that the turbines at a coal-powered power plant were upgraded, resulting in an improvement in efficiency of 3.32%. Assume that prior to the upgrade the power station had an efficiency of 36% and that the heat transfer into the engine in one day is still the same at 2.50×10^{14} J. (a) How much more electrical energy is produced due to the upgrade? (b) How much less heat transfer occurs to the environment due to the upgrade?

Solution:

(a) 8.30×10^{12} J, which is 3.32% of 2.50×10^{14} J .

(b) -8.30×10^{12} J, where the negative sign indicates a reduction in heat transfer to the environment.

Exercise:

Problem:

This problem compares the energy output and heat transfer to the environment by two different types of nuclear power stations—one with the normal efficiency of 34.0%, and another with an improved efficiency of 40.0%. Suppose both have the same heat transfer into the engine in one day, 2.50×10^{14} J. (a) How much more electrical energy is produced by the more efficient power station? (b) How much less heat transfer occurs to the environment by the more efficient power station? (One type of more efficient nuclear power station, the gas-cooled reactor, has not been reliable enough to be economically feasible in spite of its greater efficiency.)

Glossary

irreversible process

any process that depends on path direction

second law of thermodynamics

heat transfer flows from a hotter to a cooler object, never the reverse, and some heat energy in any process is lost to available work in a cyclical process

cyclical process

a process in which the path returns to its original state at the end of every cycle

Otto cycle

a thermodynamic cycle, consisting of a pair of adiabatic processes and a pair of isochoric processes, that converts heat into work, e.g., the four-stroke engine cycle of intake, compression, ignition, and exhaust

Carnot's Perfect Heat Engine: The Second Law of Thermodynamics Restated

- Identify a Carnot cycle.
- Calculate maximum theoretical efficiency of a nuclear reactor.
- Explain how dissipative processes affect the ideal Carnot engine.



This novelty toy, known as the drinking bird, is an example of Carnot's engine. It contains methylene chloride (mixed with a dye) in the abdomen, which boils at a very low temperature—about 100°F. To operate, one gets the bird's head wet. As the water evaporates, fluid moves up into the head, causing the bird to become top-heavy and dip forward back into the water. This cools down the methylene chloride in the head, and it moves back into the abdomen, causing the bird to become bottom heavy and tip up. Except for a very small input of energy—the original head-wetting—the bird becomes a perpetual motion machine of sorts. (credit: Arabesk.nl, Wikimedia Commons)

We know from the second law of thermodynamics that a heat engine cannot be 100% efficient, since there must always be some heat transfer Q_c to the environment, which is often called waste heat. How efficient, then, can a heat engine be? This question was answered at a theoretical level in 1824 by a young French engineer, Sadi Carnot (1796–1832), in his study of the then-emerging heat engine technology crucial to the Industrial Revolution. He devised a theoretical cycle, now called the **Carnot cycle**, which is the most efficient cyclical process possible. The second law of thermodynamics can be restated in terms of the Carnot cycle, and so what Carnot actually discovered was this fundamental law. Any heat engine employing the Carnot cycle is called a **Carnot engine**.

What is crucial to the Carnot cycle—and, in fact, defines it—is that only reversible processes are used. Irreversible processes involve dissipative factors, such as friction and turbulence. This increases heat transfer Q_c to the environment and reduces the efficiency of the engine. Obviously, then, reversible processes are superior.

Note:

Carnot Engine

Stated in terms of reversible processes, the **second law of thermodynamics** has a third form:

A Carnot engine operating between two given temperatures has the greatest possible efficiency of any heat engine operating between these two temperatures. Furthermore, all engines employing only reversible processes have this same maximum efficiency when operating between the same given temperatures.

[\[link\]](#) shows the PV diagram for a Carnot cycle. The cycle comprises two isothermal and two adiabatic processes. Recall that both isothermal and adiabatic processes are, in principle, reversible.

Carnot also determined the efficiency of a perfect heat engine—that is, a Carnot engine. It is always true that the efficiency of a cyclical heat engine is given by:

Equation:

$$\text{Eff} = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h}.$$

What Carnot found was that for a perfect heat engine, the ratio Q_c/Q_h equals the ratio of the absolute temperatures of the heat reservoirs. That is, $Q_c/Q_h = T_c/T_h$ for a Carnot engine, so that the maximum or **Carnot efficiency** Eff_C is given by

Equation:

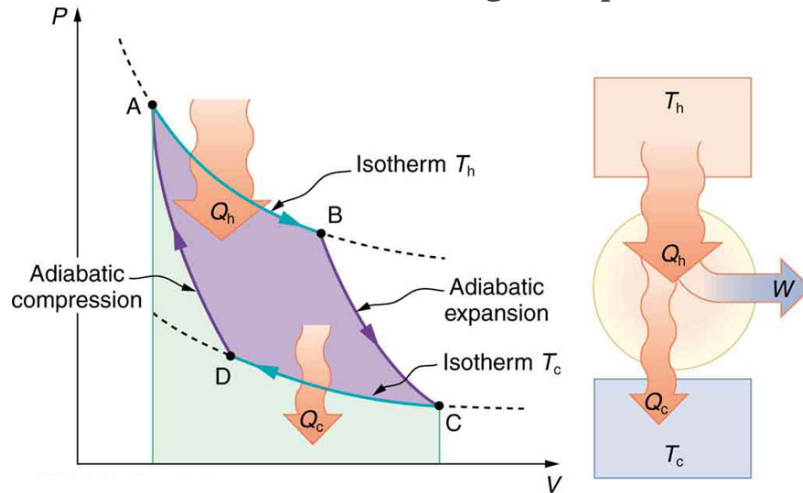
$$\text{Eff}_C = 1 - \frac{T_c}{T_h},$$

where T_h and T_c are in kelvins (or any other absolute temperature scale). No real heat engine can do as well as the Carnot efficiency—an actual efficiency of about 0.7 of this maximum is usually the best that can be accomplished. But the ideal Carnot engine, like the drinking bird above, while a fascinating novelty, has zero power. This makes it unrealistic for any applications.

Carnot's interesting result implies that 100% efficiency would be possible only if $T_c = 0$ K—that is, only if the cold reservoir were at absolute zero, a practical and theoretical impossibility. But the physical implication is this—the only way to have all heat transfer go into doing work is to remove *all* thermal energy, and this requires a cold reservoir at absolute zero.

It is also apparent that the greatest efficiencies are obtained when the ratio T_c/T_h is as small as possible. Just as discussed for the Otto cycle in the previous section, this means that efficiency is greatest for the highest possible temperature of the hot reservoir and lowest possible temperature of the cold reservoir. (This setup increases the area inside the closed loop on the PV diagram; also, it seems reasonable that the greater the temperature

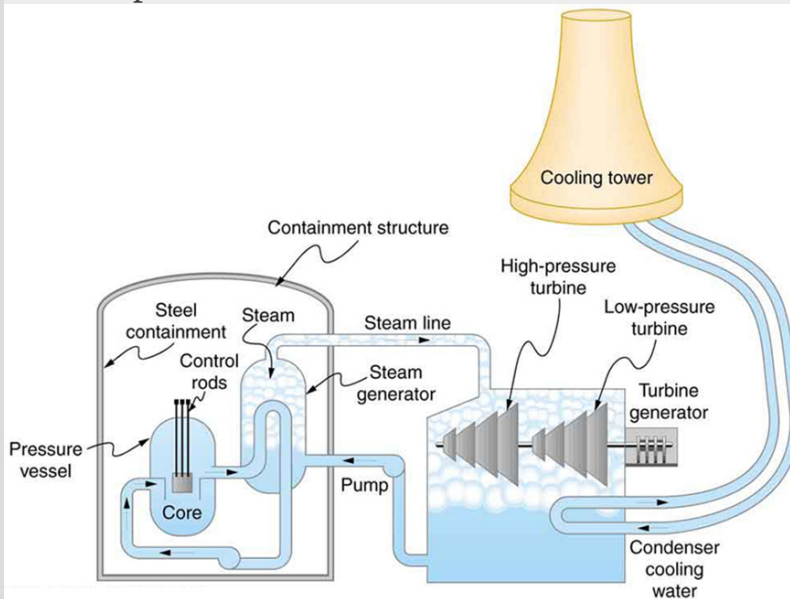
difference, the easier it is to divert the heat transfer to work.) The actual reservoir temperatures of a heat engine are usually related to the type of heat source and the temperature of the environment into which heat transfer occurs. Consider the following example.



PV diagram for a Carnot cycle, employing only reversible isothermal and adiabatic processes. Heat transfer Q_h occurs into the working substance during the isothermal path AB, which takes place at constant temperature T_h . Heat transfer Q_c occurs out of the working substance during the isothermal path CD, which takes place at constant temperature T_c . The net work output W equals the area inside the path ABCDA. Also shown is a schematic of a Carnot engine operating between hot and cold reservoirs at temperatures T_h and T_c . Any heat engine using reversible processes and operating between these two temperatures will have the same maximum efficiency as the Carnot engine.

Example:**Maximum Theoretical Efficiency for a Nuclear Reactor**

A nuclear power reactor has pressurized water at 300°C . (Higher temperatures are theoretically possible but practically not, due to limitations with materials used in the reactor.) Heat transfer from this water is a complex process (see [\[link\]](#)). Steam, produced in the steam generator, is used to drive the turbine-generators. Eventually the steam is condensed to water at 27°C and then heated again to start the cycle over. Calculate the maximum theoretical efficiency for a heat engine operating between these two temperatures.



Schematic diagram of a pressurized water nuclear reactor and the steam turbines that convert work into electrical energy. Heat exchange is used to generate steam, in part to avoid contamination of the generators with radioactivity. Two turbines are used because this is less expensive than operating a single generator that produces the same amount of electrical energy. The steam is condensed to liquid before being returned to the heat exchanger, to keep exit steam pressure low and aid the flow of steam through the turbines (equivalent to using a

lower-temperature cold reservoir). The considerable energy associated with condensation must be dissipated into the local environment; in this example, a cooling tower is used so there is no direct heat transfer to an aquatic environment. (Note that the water going to the cooling tower does not come into contact with the steam flowing over the turbines.)

Strategy

Since temperatures are given for the hot and cold reservoirs of this heat engine, $Eff_C = 1 - \frac{T_c}{T_h}$ can be used to calculate the Carnot (maximum theoretical) efficiency. Those temperatures must first be converted to kelvins.

Solution

The hot and cold reservoir temperatures are given as 300°C and 27.0°C , respectively. In kelvins, then, $T_h = 573\text{ K}$ and $T_c = 300\text{ K}$, so that the maximum efficiency is

Equation:

$$Eff_C = 1 - \frac{T_c}{T_h}.$$

Thus,

Equation:

$$\begin{aligned} Eff_C &= 1 - \frac{300\text{ K}}{573\text{ K}} \\ &= 0.476, \text{ or } 47.6\%. \end{aligned}$$

Discussion

A typical nuclear power station's actual efficiency is about 35%, a little better than 0.7 times the maximum possible value, a tribute to superior engineering. Electrical power stations fired by coal, oil, and natural gas have greater actual efficiencies (about 42%), because their boilers can reach higher temperatures and pressures. The cold reservoir temperature in

any of these power stations is limited by the local environment. [\[link\]](#) shows (a) the exterior of a nuclear power station and (b) the exterior of a coal-fired power station. Both have cooling towers into which water from the condenser enters the tower near the top and is sprayed downward, cooled by evaporation.



(a)



(b)

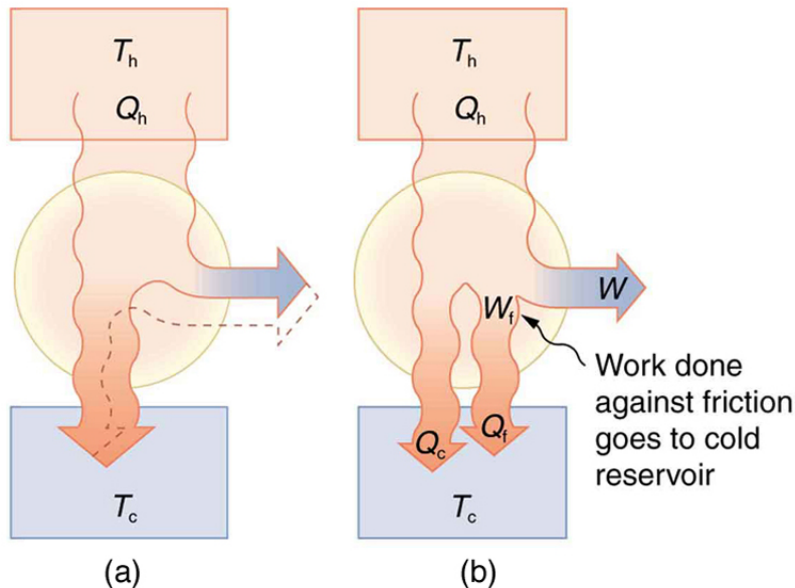
(a) A nuclear power station (credit: BlatantWorld.com) and (b) a coal-fired power station.

Both have cooling towers in which water evaporates into the environment, representing Q_c .

The nuclear reactor, which supplies Q_h , is housed inside

the dome-shaped containment buildings. (credit: Robert & Mihaela Vicol, publicphoto.org)

Since all real processes are irreversible, the actual efficiency of a heat engine can never be as great as that of a Carnot engine, as illustrated in [\[link\]](#)(a). Even with the best heat engine possible, there are always dissipative processes in peripheral equipment, such as electrical transformers or car transmissions. These further reduce the overall efficiency by converting some of the engine's work output back into heat transfer, as shown in [\[link\]](#)(b).



Real heat engines are less efficient than Carnot engines. (a) Real engines use irreversible processes, reducing the heat transfer to work. Solid lines represent the actual process; the dashed lines are what a Carnot engine would do between the same two reservoirs. (b) Friction and other dissipative processes in the output mechanisms of a heat engine convert some

of its work output into heat transfer to the environment.

Section Summary

- The Carnot cycle is a theoretical cycle that is the most efficient cyclical process possible. Any engine using the Carnot cycle, which uses only reversible processes (adiabatic and isothermal), is known as a Carnot engine.
- Any engine that uses the Carnot cycle enjoys the maximum theoretical efficiency.
- While Carnot engines are ideal engines, in reality, no engine achieves Carnot's theoretical maximum efficiency, since dissipative processes, such as friction, play a role. Carnot cycles without heat loss may be possible at absolute zero, but this has never been seen in nature.

Conceptual Questions

Exercise:

Problem:

Think about the drinking bird at the beginning of this section ([\[link\]](#)). Although the bird enjoys the theoretical maximum efficiency possible, if left to its own devices over time, the bird will cease “drinking.” What are some of the dissipative processes that might cause the bird's motion to cease?

Exercise:

Problem:

Can improved engineering and materials be employed in heat engines to reduce heat transfer into the environment? Can they eliminate heat transfer into the environment entirely?

Exercise:

Problem:

Does the second law of thermodynamics alter the conservation of energy principle?

Problem Exercises**Exercise:****Problem:**

A certain gasoline engine has an efficiency of 30.0%. What would the hot reservoir temperature be for a Carnot engine having that efficiency, if it operates with a cold reservoir temperature of 200°C?

Solution:

403°C

Exercise:**Problem:**

A gas-cooled nuclear reactor operates between hot and cold reservoir temperatures of 700°C and 27.0°C. (a) What is the maximum efficiency of a heat engine operating between these temperatures? (b) Find the ratio of this efficiency to the Carnot efficiency of a standard nuclear reactor (found in [\[link\]](#)).

Exercise:**Problem:**

(a) What is the hot reservoir temperature of a Carnot engine that has an efficiency of 42.0% and a cold reservoir temperature of 27.0°C? (b) What must the hot reservoir temperature be for a real heat engine that achieves 0.700 of the maximum efficiency, but still has an efficiency of 42.0% (and a cold reservoir at 27.0°C)? (c) Does your answer imply practical limits to the efficiency of car gasoline engines?

Solution:

(a) 244°C

(b) 477°C

(c) Yes, since automobiles engines cannot get too hot without overheating, their efficiency is limited.

Exercise:**Problem:**

Steam locomotives have an efficiency of 17.0% and operate with a hot steam temperature of 425°C. (a) What would the cold reservoir temperature be if this were a Carnot engine? (b) What would the maximum efficiency of this steam engine be if its cold reservoir temperature were 150°C?

Exercise:**Problem:**

Practical steam engines utilize 450°C steam, which is later exhausted at 270°C. (a) What is the maximum efficiency that such a heat engine can have? (b) Since 270°C steam is still quite hot, a second steam engine is sometimes operated using the exhaust of the first. What is the maximum efficiency of the second engine if its exhaust has a temperature of 150°C? (c) What is the overall efficiency of the two engines? (d) Show that this is the same efficiency as a single Carnot engine operating between 450°C and 150°C. Explicitly show how you follow the steps in the [Problem-Solving Strategies for Thermodynamics](#).

Solution:

$$(a) \text{Eff}_1 = 1 - \frac{T_{c,1}}{T_{h,1}} = 1 - \frac{543 \text{ K}}{723 \text{ K}} = 0.249 \text{ or } 24.9\%$$

$$(b) \text{Eff}_2 = 1 - \frac{423 \text{ K}}{543 \text{ K}} = 0.221 \text{ or } 22.1\%$$

$$(c) \text{Eff}_1 = 1 - \frac{T_{c,1}}{T_{h,1}} \Rightarrow T_{c,1} = T_{h,1}(1 - \text{eff}_1)$$

$$\text{similarly, } T_{c,2} = T_{h,2}(1 - \text{Eff}_2)$$

using $T_{h,2} = T_{c,1}$ in above equation gives

$$T_{c,2} = T_{h,1}(1 - \text{Eff}_1)(1 - \text{Eff}_2) \equiv T_{h,1}(1 - \text{Eff}_{\text{overall}})$$

$$\therefore (1 - \text{Eff}_{\text{overall}}) = (1 - \text{Eff}_1)(1 - \text{Eff}_2)$$

$$\text{Eff}_{\text{overall}} = 1 - (1 - 0.249)(1 - 0.221) = 41.5\%$$

$$(d) \text{Eff}_{\text{overall}} = 1 - \frac{423 \text{ K}}{723 \text{ K}} = 0.415 \text{ or } 41.5\%$$

Exercise:

Problem:

A coal-fired electrical power station has an efficiency of 38%. The temperature of the steam leaving the boiler is 550°C. What percentage of the maximum efficiency does this station obtain? (Assume the temperature of the environment is 20°C.)

Exercise:

Problem:

Would you be willing to financially back an inventor who is marketing a device that she claims has 25 kJ of heat transfer at 600 K, has heat transfer to the environment at 300 K, and does 12 kJ of work? Explain your answer.

Solution:

The heat transfer to the cold reservoir is

$$Q_c = Q_h - W = 25 \text{ kJ} - 12 \text{ kJ} = 13 \text{ kJ}, \text{ so the efficiency is}$$

$$\text{Eff} = 1 - \frac{Q_c}{Q_h} = 1 - \frac{13 \text{ kJ}}{25 \text{ kJ}} = 0.48. \text{ The Carnot efficiency is}$$

$$\text{Eff}_C = 1 - \frac{T_c}{T_h} = 1 - \frac{300 \text{ K}}{600 \text{ K}} = 0.50. \text{ The actual efficiency is } 96\%$$

of the Carnot efficiency, which is much higher than the best-ever achieved of about 70%, so her scheme is likely to be fraudulent.

Exercise:

Problem: Unreasonable Results

(a) Suppose you want to design a steam engine that has heat transfer to the environment at 270°C and has a Carnot efficiency of 0.800. What temperature of hot steam must you use? (b) What is unreasonable about the temperature? (c) Which premise is unreasonable?

Exercise:

Problem: Unreasonable Results

Calculate the cold reservoir temperature of a steam engine that uses hot steam at 450°C and has a Carnot efficiency of 0.700. (b) What is unreasonable about the temperature? (c) Which premise is unreasonable?

Solution:

(a) -56.3°C

(b) The temperature is too cold for the output of a steam engine (the local environment). It is below the freezing point of water.

(c) The assumed efficiency is too high.

Glossary

Carnot cycle

a cyclical process that uses only reversible processes, the adiabatic and isothermal processes

Carnot engine

a heat engine that uses a Carnot cycle

Carnot efficiency

the maximum theoretical efficiency for a heat engine

Applications of Thermodynamics: Heat Pumps and Refrigerators

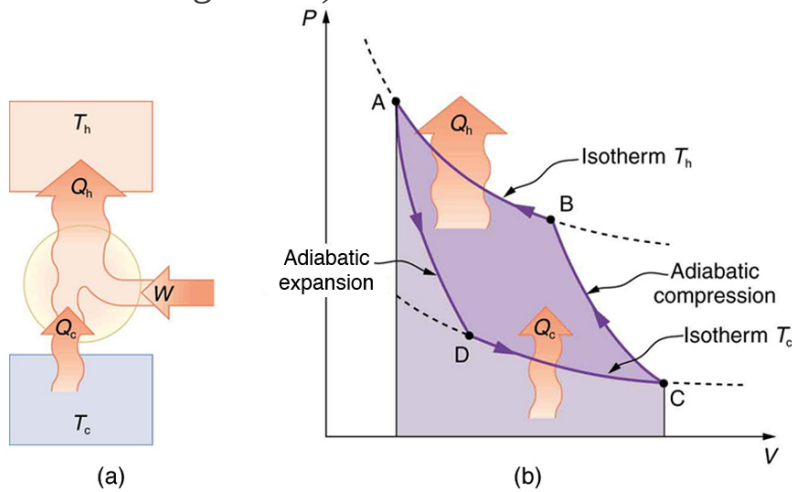
- Describe the use of heat engines in heat pumps and refrigerators.
- Demonstrate how a heat pump works to warm an interior space.
- Explain the differences between heat pumps and refrigerators.
- Calculate a heat pump's coefficient of performance.



Almost every home contains a refrigerator. Most people don't realize they are also sharing their homes with a heat pump.
(credit: Id1337x, Wikimedia Commons)

Heat pumps, air conditioners, and refrigerators utilize heat transfer from cold to hot. They are heat engines run backward. We say backward, rather than reverse, because except for Carnot engines, all heat engines, though they can be run backward, cannot truly be reversed. Heat transfer occurs from a cold reservoir Q_c and into a hot one. This requires work input W , which is also converted to heat transfer. Thus the heat transfer to the hot reservoir is $Q_h = Q_c + W$. (Note that Q_h , Q_c , and W are positive, with their directions indicated on schematics rather than by sign.) A heat pump's mission is for heat transfer Q_h to occur into a warm environment, such as a home in the winter. The mission of air conditioners and refrigerators is for

heat transfer Q_c to occur from a cool environment, such as chilling a room or keeping food at lower temperatures than the environment. (Actually, a heat pump can be used both to heat and cool a space. It is essentially an air conditioner and a heating unit all in one. In this section we will concentrate on its heating mode.)

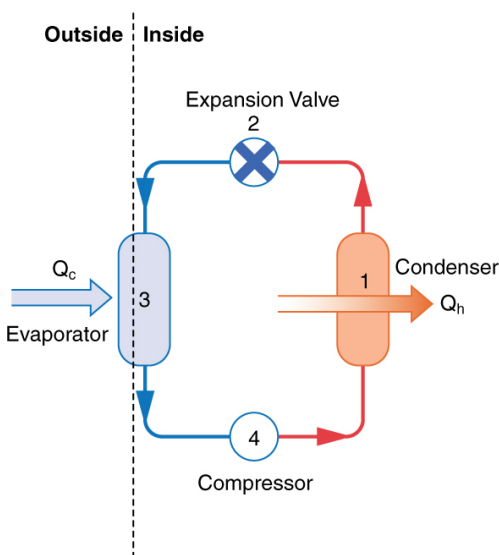


Heat pumps, air conditioners, and refrigerators are heat engines operated backward. The one shown here is based on a Carnot (reversible) engine. (a) Schematic diagram showing heat transfer from a cold reservoir to a warm reservoir with a heat pump. The directions of W , Q_h , and Q_c are opposite what they would be in a heat engine. (b) PV diagram for a Carnot cycle similar to that in [\[link\]](#) but reversed, following path ADCBA. The area inside the loop is negative, meaning there is a net work input. There is heat transfer Q_c into the system from a cold reservoir along path DC, and heat transfer Q_h out of the system into a hot reservoir along path BA.

Heat Pumps

The great advantage of using a heat pump to keep your home warm, rather than just burning fuel, is that a heat pump supplies $Q_h = Q_c + W$. Heat transfer is from the outside air, even at a temperature below freezing, to the indoor space. You only pay for W , and you get an additional heat transfer of Q_c from the outside at no cost; in many cases, at least twice as much energy is transferred to the heated space as is used to run the heat pump. When you burn fuel to keep warm, you pay for all of it. The disadvantage is that the work input (required by the second law of thermodynamics) is sometimes more expensive than simply burning fuel, especially if the work is done by electrical energy.

The basic components of a heat pump in its heating mode are shown in [\[link\]](#). A working fluid such as a non-CFC refrigerant is used. In the outdoor coils (the evaporator), heat transfer Q_c occurs to the working fluid from the cold outdoor air, turning it into a gas.



A simple heat pump has four basic components:

- (1) condenser,
- (2) expansion valve,
- (3) evaporator, and
- (4) compressor. In the

heating mode, heat transfer Q_c occurs to the working fluid in the evaporator (3) from the colder outdoor air, turning it into a gas. The electrically driven compressor (4) increases the temperature and pressure of the gas and forces it into the condenser coils (1) inside the heated space. Because the temperature of the gas is higher than the temperature in the room, heat transfer from the gas to the room occurs as the gas condenses to a liquid. The working fluid is then cooled as it flows back through an expansion valve (2) to the outdoor evaporator coils.

The electrically driven compressor (work input W) raises the temperature and pressure of the gas and forces it into the condenser coils that are inside the heated space. Because the temperature of the gas is higher than the temperature inside the room, heat transfer to the room occurs and the gas condenses to a liquid. The liquid then flows back through a pressure-reducing valve to the outdoor evaporator coils, being cooled through expansion. (In a cooling cycle, the evaporator and condenser coils exchange roles and the flow direction of the fluid is reversed.)

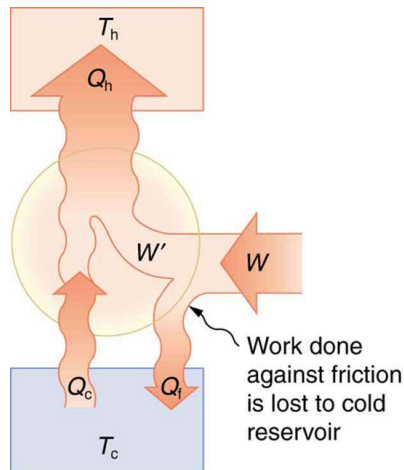
The quality of a heat pump is judged by how much heat transfer Q_h occurs into the warm space compared with how much work input W is required. In the spirit of taking the ratio of what you get to what you spend, we define a **heat pump's coefficient of performance** (COP_{hp}) to be

Equation:

$$COP_{hp} = \frac{Q_h}{W}.$$

Since the efficiency of a heat engine is $Eff = W/Q_h$, we see that $COP_{hp} = 1/Eff$, an important and interesting fact. First, since the efficiency of any heat engine is less than 1, it means that COP_{hp} is always greater than 1—that is, a heat pump always has more heat transfer Q_h than work put into it. Second, it means that heat pumps work best when temperature differences are small. The efficiency of a perfect, or Carnot, engine is $Eff_C = 1 - (T_c/T_h)$; thus, the smaller the temperature difference, the smaller the efficiency and the greater the COP_{hp} (because $COP_{hp} = 1/Eff$). In other words, heat pumps do not work as well in very cold climates as they do in more moderate climates.

Friction and other irreversible processes reduce heat engine efficiency, but they do *not* benefit the operation of a heat pump—instead, they reduce the work input by converting part of it to heat transfer back into the cold reservoir before it gets into the heat pump.



When a real heat engine is run backward, some of the intended work input (W) goes into heat transfer before it gets into the heat engine, thereby reducing its coefficient of performance COP_{hp} . In this figure, W' represents the portion of W that goes into the heat pump, while the remainder of W is lost in the form of frictional heat (Q_f) to the cold reservoir. If all of W had gone into the heat pump, then Q_h would have

been greater. The best heat pump uses adiabatic and isothermal processes, since, in theory, there would be no dissipative processes to reduce the heat transfer to the hot reservoir.

Example:

The Best COP_{hp} of a Heat Pump for Home Use

A heat pump used to warm a home must employ a cycle that produces a working fluid at temperatures greater than typical indoor temperature so that heat transfer to the inside can take place. Similarly, it must produce a working fluid at temperatures that are colder than the outdoor temperature so that heat transfer occurs from outside. Its hot and cold reservoir temperatures therefore cannot be too close, placing a limit on its COP_{hp} . (See [\[link\]](#).) What is the best coefficient of performance possible for such a heat pump, if it has a hot reservoir temperature of 45.0°C and a cold reservoir temperature of -15.0°C ?

Strategy

A Carnot engine reversed will give the best possible performance as a heat pump. As noted above, $COP_{hp} = 1/Eff$, so that we need to first calculate the Carnot efficiency to solve this problem.

Solution

Carnot efficiency in terms of absolute temperature is given by:

Equation:

$$Eff_C = 1 - \frac{T_c}{T_h}.$$

The temperatures in kelvins are $T_h = 318 \text{ K}$ and $T_c = 258 \text{ K}$, so that

Equation:

$$Eff_C = 1 - \frac{258 \text{ K}}{318 \text{ K}} = 0.1887.$$

Thus, from the discussion above,

Equation:

$$COP_{hp} = \frac{1}{Eff} = \frac{1}{0.1887} = 5.30,$$

or

Equation:

$$COP_{hp} = \frac{Q_h}{W} = 5.30,$$

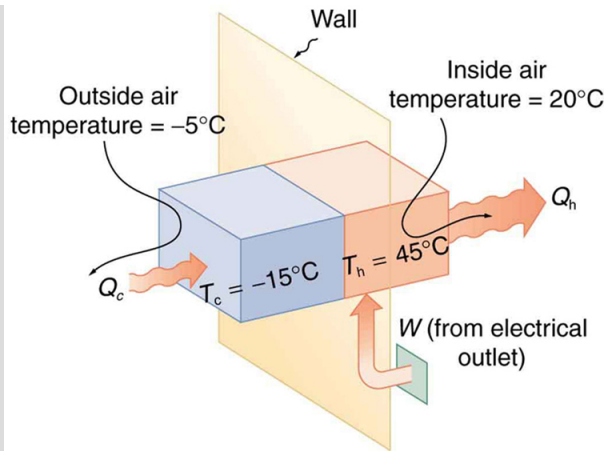
so that

Equation:

$$Q_h = 5.30 \text{ W}.$$

Discussion

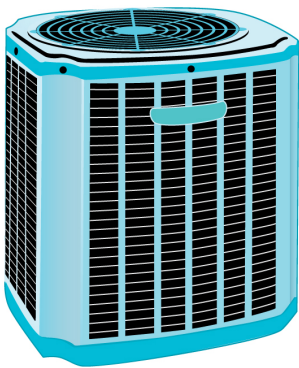
This result means that the heat transfer by the heat pump is 5.30 times as much as the work put into it. It would cost 5.30 times as much for the same heat transfer by an electric room heater as it does for that produced by this heat pump. This is not a violation of conservation of energy. Cold ambient air provides 4.3 J per 1 J of work from the electrical outlet.



Heat transfer from the outside to the inside, along with work done to run the pump, takes place in the heat pump of the example above. Note that the cold temperature produced by the heat pump is lower than the outside temperature, so that heat transfer into the working fluid occurs. The pump's compressor produces a temperature greater than the indoor temperature in order for heat transfer into the house to occur.

Real heat pumps do not perform quite as well as the ideal one in the previous example; their values of COP_{hp} range from about 2 to 4. This range means that the heat transfer Q_h from the heat pumps is 2 to 4 times as great as the work W put into them. Their economical feasibility is still limited, however, since W is usually supplied by electrical energy that costs more per joule than heat transfer by burning fuels like natural gas. Furthermore, the initial cost of a heat pump is greater than that of many

furnaces, so that a heat pump must last longer for its cost to be recovered. Heat pumps are most likely to be economically superior where winter temperatures are mild, electricity is relatively cheap, and other fuels are relatively expensive. Also, since they can cool as well as heat a space, they have advantages where cooling in summer months is also desired. Thus some of the best locations for heat pumps are in warm summer climates with cool winters. [\[link\]](#) shows a heat pump, called a “*reverse cycle*” or “*split-system cooler*” in some countries.



In hot weather, heat transfer occurs from air inside the room to air outside, cooling the room. In cool weather, heat transfer occurs from air outside to air inside, warming the room. This switching is

achieved by
reversing the
direction of
flow of the
working
fluid.

Air Conditioners and Refrigerators

Air conditioners and refrigerators are designed to cool something down in a warm environment. As with heat pumps, work input is required for heat transfer from cold to hot, and this is expensive. The quality of air conditioners and refrigerators is judged by how much heat transfer Q_c occurs from a cold environment compared with how much work input W is required. What is considered the benefit in a heat pump is considered waste heat in a refrigerator. We thus define the **coefficient of performance** (COP_{ref}) of an air conditioner or refrigerator to be

Equation:

$$COP_{\text{ref}} = \frac{Q_c}{W}.$$

Noting again that $Q_h = Q_c + W$, we can see that an air conditioner will have a lower coefficient of performance than a heat pump, because $COP_{\text{hp}} = Q_h/W$ and Q_h is greater than Q_c . In this module's Problems and Exercises, you will show that

Equation:

$$COP_{\text{ref}} = COP_{\text{hp}} - 1$$

for a heat engine used as either an air conditioner or a heat pump operating between the same two temperatures. Real air conditioners and refrigerators typically do remarkably well, having values of COP_{ref} ranging from 2 to 6.

These numbers are better than the COP_{hp} values for the heat pumps mentioned above, because the temperature differences are smaller, but they are less than those for Carnot engines operating between the same two temperatures.

A type of COP rating system called the “energy efficiency rating” (EER) has been developed. This rating is an example where non-SI units are still used and relevant to consumers. To make it easier for the consumer, Australia, Canada, New Zealand, and the U.S. use an Energy Star Rating out of 5 stars—the more stars, the more energy efficient the appliance. $EERs$ are expressed in mixed units of British thermal units (Btu) per hour of heating or cooling divided by the power input in watts. Room air conditioners are readily available with $EERs$ ranging from 6 to 12. Although not the same as the $COPs$ just described, these $EERs$ are good for comparison purposes—the greater the EER , the cheaper an air conditioner is to operate (but the higher its purchase price is likely to be).

The EER of an air conditioner or refrigerator can be expressed as

Equation:

$$EER = \frac{Q_c/t_1}{W/t_2},$$

where Q_c is the amount of heat transfer from a cold environment in British thermal units, t_1 is time in hours, W is the work input in joules, and t_2 is time in seconds.

Note:

Problem-Solving Strategies for Thermodynamics

1. *Examine the situation to determine whether heat, work, or internal energy are involved.* Look for any system where the primary methods of transferring energy are heat and work. Heat engines, heat pumps, refrigerators, and air conditioners are examples of such systems.

2. *Identify the system of interest and draw a labeled diagram of the system showing energy flow.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful. Maximum efficiency means a Carnot engine is involved. Efficiency is not the same as the coefficient of performance.*
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns). Be sure to distinguish heat transfer into a system from heat transfer out of the system, as well as work input from work output. In many situations, it is useful to determine the type of process, such as isothermal or adiabatic.*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).*
6. *Substitute the known quantities along with their units into the appropriate equation and obtain numerical solutions complete with units.*
7. *Check the answer to see if it is reasonable: Does it make sense? For example, efficiency is always less than 1, whereas coefficients of performance are greater than 1.*

Section Summary

- An artifact of the second law of thermodynamics is the ability to heat an interior space using a heat pump. Heat pumps compress cold ambient air and, in so doing, heat it to room temperature without violation of conservation principles.
- To calculate the heat pump's coefficient of performance, use the equation $COP_{hp} = \frac{Q_h}{W}$.
- A refrigerator is a heat pump; it takes warm ambient air and expands it to chill it.

Conceptual Questions

Exercise:

Problem:

Explain why heat pumps do not work as well in very cold climates as they do in milder ones. Is the same true of refrigerators?

Exercise:**Problem:**

In some Northern European nations, homes are being built without heating systems of any type. They are very well insulated and are kept warm by the body heat of the residents. However, when the residents are not at home, it is still warm in these houses. What is a possible explanation?

Exercise:**Problem:**

Why do refrigerators, air conditioners, and heat pumps operate most cost-effectively for cycles with a small difference between T_h and T_c ? (Note that the temperatures of the cycle employed are crucial to its COP.)

Exercise:**Problem:**

Grocery store managers contend that there is *less* total energy consumption in the summer if the store is kept at a *low* temperature. Make arguments to support or refute this claim, taking into account that there are numerous refrigerators and freezers in the store.

Exercise:**Problem:**

Can you cool a kitchen by leaving the refrigerator door open?

Problem Exercises

Exercise:**Problem:**

What is the coefficient of performance of an ideal heat pump that has heat transfer from a cold temperature of -25.0°C to a hot temperature of 40.0°C ?

Solution:

4.82

Exercise:**Problem:**

Suppose you have an ideal refrigerator that cools an environment at -20.0°C and has heat transfer to another environment at 50.0°C . What is its coefficient of performance?

Exercise:**Problem:**

What is the best coefficient of performance possible for a hypothetical refrigerator that could make liquid nitrogen at -200°C and has heat transfer to the environment at 35.0°C ?

Solution:

0.311

Exercise:**Problem:**

In a very mild winter climate, a heat pump has heat transfer from an environment at 5.00°C to one at 35.0°C . What is the best possible coefficient of performance for these temperatures? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Thermodynamics](#).

Exercise:**Problem:**

(a) What is the best coefficient of performance for a heat pump that has a hot reservoir temperature of 50.0°C and a cold reservoir temperature of -20.0°C ? (b) How much heat transfer occurs into the warm environment if $3.60 \times 10^7 \text{ J}$ of work ($10.0 \text{ kW} \cdot \text{h}$) is put into it? (c) If the cost of this work input is 10.0 cents/ $\text{kW} \cdot \text{h}$, how does its cost compare with the direct heat transfer achieved by burning natural gas at a cost of 85.0 cents per therm. (A therm is a common unit of energy for natural gas and equals $1.055 \times 10^8 \text{ J}$.)

Solution:

(a) 4.61

(b) $1.66 \times 10^8 \text{ J}$ or $3.97 \times 10^4 \text{ kcal}$

(c) To transfer $1.66 \times 10^8 \text{ J}$, heat pump costs \$1.00, natural gas costs \$1.34.

Exercise:**Problem:**

(a) What is the best coefficient of performance for a refrigerator that cools an environment at -30.0°C and has heat transfer to another environment at 45.0°C ? (b) How much work in joules must be done for a heat transfer of 4186 kJ from the cold environment? (c) What is the cost of doing this if the work costs 10.0 cents per $3.60 \times 10^6 \text{ J}$ (a kilowatt-hour)? (d) How many kJ of heat transfer occurs into the warm environment? (e) Discuss what type of refrigerator might operate between these temperatures.

Exercise:

Problem:

Suppose you want to operate an ideal refrigerator with a cold temperature of -10.0°C , and you would like it to have a coefficient of performance of 7.00. What is the hot reservoir temperature for such a refrigerator?

Solution:

27.6°C

Exercise:**Problem:**

An ideal heat pump is being considered for use in heating an environment with a temperature of 22.0°C . What is the cold reservoir temperature if the pump is to have a coefficient of performance of 12.0?

Exercise:**Problem:**

A 4-ton air conditioner removes 5.06×10^7 J (48,000 British thermal units) from a cold environment in 1.00 h. (a) What energy input in joules is necessary to do this if the air conditioner has an energy efficiency rating (*EER*) of 12.0? (b) What is the cost of doing this if the work costs 10.0 cents per 3.60×10^6 J (one kilowatt-hour)? (c) Discuss whether this cost seems realistic. Note that the energy efficiency rating (*EER*) of an air conditioner or refrigerator is defined to be the number of British thermal units of heat transfer from a cold environment per hour divided by the watts of power input.

Solution:

(a) 1.44×10^7 J

(b) 40 cents

(c) This cost seems quite realistic; it says that running an air conditioner all day would cost \$9.59 (if it ran continuously).

Exercise:

Problem:

Show that the coefficients of performance of refrigerators and heat pumps are related by $COP_{\text{ref}} = COP_{\text{hp}} - 1$.

Start with the definitions of the COP s and the conservation of energy relationship between Q_h , Q_c , and W .

Glossary

heat pump

a machine that generates heat transfer from cold to hot

coefficient of performance

for a heat pump, it is the ratio of heat transfer at the output (the hot reservoir) to the work supplied; for a refrigerator or air conditioner, it is the ratio of heat transfer from the cold reservoir to the work supplied

Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy

- Define entropy and calculate the increase of entropy in a system with reversible and irreversible processes.
- Explain the expected fate of the universe in entropic terms.
- Calculate the increasing disorder of a system.



The ice in this drink is slowly melting. Eventually the liquid will reach thermal equilibrium, as predicted by the second law of thermodynamics.

(credit: Jon Sullivan, PDPhoto.org)

There is yet another way of expressing the second law of thermodynamics. This version relates to a concept called **entropy**. By examining it, we shall

see that the directions associated with the second law—heat transfer from hot to cold, for example—are related to the tendency in nature for systems to become disordered and for less energy to be available for use as work. The entropy of a system can in fact be shown to be a measure of its disorder and of the unavailability of energy to do work.

Note:

Making Connections: Entropy, Energy, and Work

Recall that the simple definition of energy is the ability to do work. Entropy is a measure of how much energy is not available to do work. Although all forms of energy are interconvertible, and all can be used to do work, it is not always possible, even in principle, to convert the entire available energy into work. That unavailable energy is of interest in thermodynamics, because the field of thermodynamics arose from efforts to convert heat to work.

We can see how entropy is defined by recalling our discussion of the Carnot engine. We noted that for a Carnot cycle, and hence for any reversible processes, $Q_c/Q_h = T_c/T_h$. Rearranging terms yields

Equation:

$$\frac{Q_c}{T_c} = \frac{Q_h}{T_h}$$

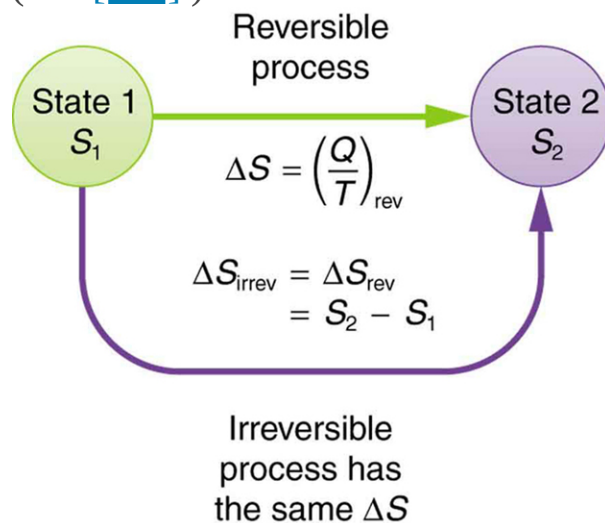
for any reversible process. Q_c and Q_h are absolute values of the heat transfer at temperatures T_c and T_h , respectively. This ratio of Q/T is defined to be the **change in entropy** ΔS for a reversible process,

Equation:

$$\Delta S = \left(\frac{Q}{T} \right)_{\text{rev}},$$

where Q is the heat transfer, which is positive for heat transfer into and negative for heat transfer out of, and T is the absolute temperature at which the reversible process takes place. The SI unit for entropy is joules per kelvin (J/K). If temperature changes during the process, then it is usually a good approximation (for small changes in temperature) to take T to be the average temperature, avoiding the need to use integral calculus to find ΔS .

The definition of ΔS is strictly valid only for reversible processes, such as used in a Carnot engine. However, we can find ΔS precisely even for real, irreversible processes. The reason is that the entropy S of a system, like internal energy U , depends only on the state of the system and not how it reached that condition. Entropy is a property of state. Thus the change in entropy ΔS of a system between state 1 and state 2 is the same no matter how the change occurs. We just need to find or imagine a reversible process that takes us from state 1 to state 2 and calculate ΔS for that process. That will be the change in entropy for any process going from state 1 to state 2. (See [\[link\]](#).)



When a system goes from state 1 to state 2, its entropy changes by the same amount ΔS , whether a hypothetical reversible path is followed or a real irreversible path is taken.

Now let us take a look at the change in entropy of a Carnot engine and its heat reservoirs for one full cycle. The hot reservoir has a loss of entropy $\Delta S_h = -Q_h/T_h$, because heat transfer occurs out of it (remember that when heat transfers out, then Q has a negative sign). The cold reservoir has a gain of entropy $\Delta S_c = Q_c/T_c$, because heat transfer occurs into it. (We assume the reservoirs are sufficiently large that their temperatures are constant.) So the total change in entropy is

Equation:

$$\Delta S_{\text{tot}} = \Delta S_h + \Delta S_c.$$

Thus, since we know that $Q_h/T_h = Q_c/T_c$ for a Carnot engine,

Equation:

$$\Delta S_{\text{tot}} = -\frac{Q_h}{T_h} + \frac{Q_c}{T_c} = 0.$$

This result, which has general validity, means that *the total change in entropy for a system in any reversible process is zero.*

The entropy of various parts of the system may change, but the total change is zero. Furthermore, the system does not affect the entropy of its surroundings, since heat transfer between them does not occur. Thus the reversible process changes neither the total entropy of the system nor the entropy of its surroundings. Sometimes this is stated as follows: *Reversible processes do not affect the total entropy of the universe.* Real processes are not reversible, though, and they do change total entropy. We can, however, use hypothetical reversible processes to determine the value of entropy in real, irreversible processes. The following example illustrates this point.

Example:

Entropy Increases in an Irreversible (Real) Process

Spontaneous heat transfer from hot to cold is an irreversible process. Calculate the total change in entropy if 4000 J of heat transfer occurs from

a hot reservoir at $T_h = 600 \text{ K}$ (327° C) to a cold reservoir at $T_c = 250 \text{ K}$ (-23° C), assuming there is no temperature change in either reservoir. (See [\[link\]](#).)

Strategy

How can we calculate the change in entropy for an irreversible process when $\Delta S_{\text{tot}} = \Delta S_h + \Delta S_c$ is valid only for reversible processes? Remember that the total change in entropy of the hot and cold reservoirs will be the same whether a reversible or irreversible process is involved in heat transfer from hot to cold. So we can calculate the change in entropy of the hot reservoir for a hypothetical reversible process in which 4000 J of heat transfer occurs from it; then we do the same for a hypothetical reversible process in which 4000 J of heat transfer occurs to the cold reservoir. This produces the same changes in the hot and cold reservoirs that would occur if the heat transfer were allowed to occur irreversibly between them, and so it also produces the same changes in entropy.

Solution

We now calculate the two changes in entropy using $\Delta S_{\text{tot}} = \Delta S_h + \Delta S_c$. First, for the heat transfer from the hot reservoir,

Equation:

$$\Delta S_h = \frac{-Q_h}{T_h} = \frac{-4000 \text{ J}}{600 \text{ K}} = -6.67 \text{ J/K}.$$

And for the cold reservoir,

Equation:

$$\Delta S_c = \frac{Q_c}{T_c} = \frac{4000 \text{ J}}{250 \text{ K}} = 16.0 \text{ J/K}.$$

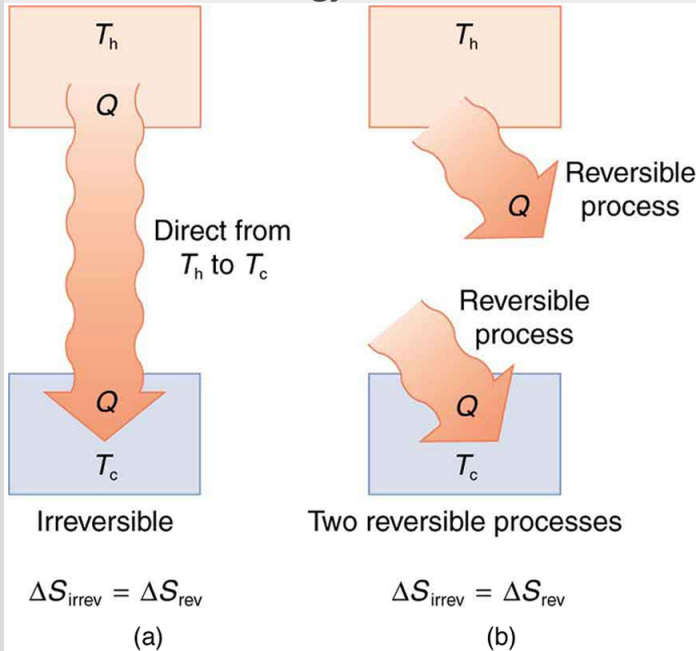
Thus the total is

Equation:

$$\begin{aligned}\Delta S_{\text{tot}} &= \Delta S_h + \Delta S_c \\ &= (-6.67 + 16.0) \text{ J/K} \\ &= 9.33 \text{ J/K}.\end{aligned}$$

Discussion

There is an *increase* in entropy for the system of two heat reservoirs undergoing this irreversible heat transfer. We will see that this means there is a loss of ability to do work with this transferred energy. Entropy has increased, and energy has become unavailable to do work.



(a) Heat transfer from a hot object to a cold one is an irreversible process that produces an overall increase in entropy. (b) The same final state and, thus, the same change in entropy is achieved for the objects if reversible heat transfer processes occur between the two objects whose temperatures are the same as the temperatures of the corresponding objects in the irreversible process.

It is reasonable that entropy increases for heat transfer from hot to cold. Since the change in entropy is Q/T , there is a larger change at lower

temperatures. The decrease in entropy of the hot object is therefore less than the increase in entropy of the cold object, producing an overall increase, just as in the previous example. This result is very general:

There is an increase in entropy for any system undergoing an irreversible process.

With respect to entropy, there are only two possibilities: entropy is constant for a reversible process, and it increases for an irreversible process. There is a fourth version of **the second law of thermodynamics stated in terms of entropy**:

The total entropy of a system either increases or remains constant in any process; it never decreases.

For example, heat transfer cannot occur spontaneously from cold to hot, because entropy would decrease.

Entropy is very different from energy. Entropy is *not* conserved but increases in all real processes. Reversible processes (such as in Carnot engines) are the processes in which the most heat transfer to work takes place and are also the ones that keep entropy constant. Thus we are led to make a connection between entropy and the availability of energy to do work.

Entropy and the Unavailability of Energy to Do Work

What does a change in entropy mean, and why should we be interested in it? One reason is that entropy is directly related to the fact that not all heat transfer can be converted into work. The next example gives some indication of how an increase in entropy results in less heat transfer into work.

Example:

Less Work is Produced by a Given Heat Transfer When Entropy Change is Greater

(a) Calculate the work output of a Carnot engine operating between temperatures of 600 K and 100 K for 4000 J of heat transfer to the engine.
(b) Now suppose that the 4000 J of heat transfer occurs first from the 600 K reservoir to a 250 K reservoir (without doing any work, and this produces the increase in entropy calculated above) before transferring into a Carnot engine operating between 250 K and 100 K. What work output is produced? (See [\[link\]](#).)

Strategy

In both parts, we must first calculate the Carnot efficiency and then the work output.

Solution (a)

The Carnot efficiency is given by

Equation:

$$Eff_C = 1 - \frac{T_c}{T_h}.$$

Substituting the given temperatures yields

Equation:

$$Eff_C = 1 - \frac{100 \text{ K}}{600 \text{ K}} = 0.833.$$

Now the work output can be calculated using the definition of efficiency for any heat engine as given by

Equation:

$$Eff = \frac{W}{Q_h}.$$

Solving for W and substituting known terms gives

Equation:

$$\begin{aligned} W &= Eff_C Q_h \\ &= (0.833)(4000 \text{ J}) = 3333 \text{ J}. \end{aligned}$$

Solution (b)

Similarly,

Equation:

$$Eff'_C = 1 - \frac{T_c}{T'_c} = 1 - \frac{100 \text{ K}}{250 \text{ K}} = 0.600,$$

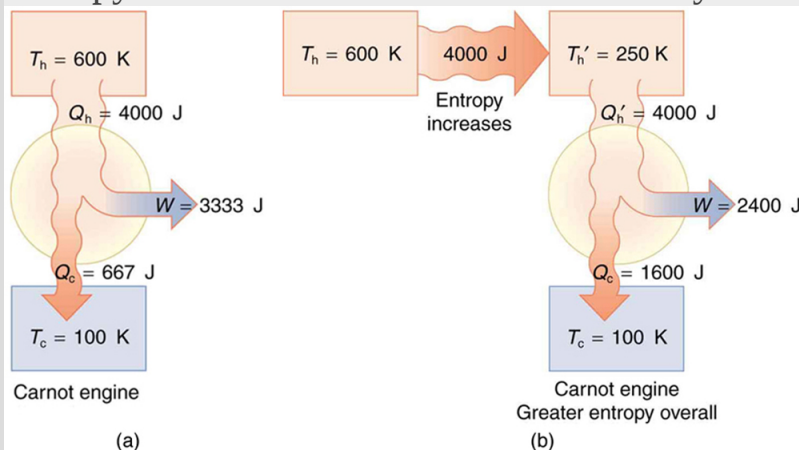
so that

Equation:

$$\begin{aligned} W &= Eff'_C Q_h \\ &= (0.600)(4000 \text{ J}) = 2400 \text{ J}. \end{aligned}$$

Discussion

There is 933 J less work from the same heat transfer in the second process. This result is important. The same heat transfer into two perfect engines produces different work outputs, because the entropy change differs in the two cases. In the second case, entropy is greater and less work is produced. Entropy is associated with the *unavailability* of energy to do work.



(a) A Carnot engine working at between 600 K and 100 K has 4000 J of heat transfer and performs 3333 J of work. (b) The 4000 J of heat transfer occurs first irreversibly to a 250 K reservoir and then goes into a Carnot engine. The increase in entropy caused by the heat transfer to a colder reservoir results in a smaller work output of 2400 J. There is a permanent loss of 933 J of energy for the purpose of doing work.

When entropy increases, a certain amount of energy becomes *permanently* unavailable to do work. The energy is not lost, but its character is changed, so that some of it can never be converted to doing work—that is, to an organized force acting through a distance. For instance, in the previous example, 933 J less work was done after an increase in entropy of 9.33 J/K occurred in the 4000 J heat transfer from the 600 K reservoir to the 250 K reservoir. It can be shown that the amount of energy that becomes unavailable for work is

Equation:

$$W_{\text{unavail}} = \Delta S \cdot T_0,$$

where T_0 is the lowest temperature utilized. In the previous example,

Equation:

$$W_{\text{unavail}} = (9.33 \text{ J/K})(100 \text{ K}) = 933 \text{ J}$$

as found.

Heat Death of the Universe: An Overdose of Entropy

In the early, energetic universe, all matter and energy were easily interchangeable and identical in nature. Gravity played a vital role in the young universe. Although it may have *seemed* disorderly, and therefore, superficially entropic, in fact, there was enormous potential energy available to do work—all the future energy in the universe.

As the universe matured, temperature differences arose, which created more opportunity for work. Stars are hotter than planets, for example, which are warmer than icy asteroids, which are warmer still than the vacuum of the space between them.

Most of these are cooling down from their usually violent births, at which time they were provided with energy of their own—nuclear energy in the case of stars, volcanic energy on Earth and other planets, and so on. Without additional energy input, however, their days are numbered.

As entropy increases, less and less energy in the universe is available to do work. On Earth, we still have great stores of energy such as fossil and nuclear fuels; large-scale temperature differences, which can provide wind energy; geothermal energies due to differences in temperature in Earth's layers; and tidal energies owing to our abundance of liquid water. As these are used, a certain fraction of the energy they contain can never be converted into doing work. Eventually, all fuels will be exhausted, all temperatures will equalize, and it will be impossible for heat engines to function, or for work to be done.

Entropy increases in a closed system, such as the universe. But in parts of the universe, for instance, in the Solar system, it is not a locally closed system. Energy flows from the Sun to the planets, replenishing Earth's stores of energy. The Sun will continue to supply us with energy for about another five billion years. We will enjoy direct solar energy, as well as side effects of solar energy, such as wind power and biomass energy from photosynthetic plants. The energy from the Sun will keep our water at the liquid state, and the Moon's gravitational pull will continue to provide tidal energy. But Earth's geothermal energy will slowly run down and won't be replenished.

But in terms of the universe, and the very long-term, very large-scale picture, the entropy of the universe is increasing, and so the availability of energy to do work is constantly decreasing. Eventually, when all stars have died, all forms of potential energy have been utilized, and all temperatures have equalized (depending on the mass of the universe, either at a very high temperature following a universal contraction, or a very low one, just before all activity ceases) there will be no possibility of doing work.

Either way, the universe is destined for thermodynamic equilibrium—maximum entropy. This is often called the *heat death of the universe*, and will mean the end of all activity. However, whether the universe contracts and heats up, or continues to expand and cools down, the end is not near.

Calculations of black holes suggest that entropy can easily continue for at least 10^{100} years.

Order to Disorder

Entropy is related not only to the unavailability of energy to do work—it is also a measure of disorder. This notion was initially postulated by Ludwig Boltzmann in the 1800s. For example, melting a block of ice means taking a highly structured and orderly system of water molecules and converting it into a disorderly liquid in which molecules have no fixed positions. (See [\[link\]](#).) There is a large increase in entropy in the process, as seen in the following example.

Example:

Entropy Associated with Disorder

Find the increase in entropy of 1.00 kg of ice originally at 0°C that is melted to form water at 0°C .

Strategy

As before, the change in entropy can be calculated from the definition of ΔS once we find the energy Q needed to melt the ice.

Solution

The change in entropy is defined as:

Equation:

$$\Delta S = \frac{Q}{T}.$$

Here Q is the heat transfer necessary to melt 1.00 kg of ice and is given by

Equation:

$$Q = mL_f,$$

where m is the mass and L_f is the latent heat of fusion. $L_f = 334\text{ kJ/kg}$ for water, so that

Equation:

$$Q = (1.00 \text{ kg})(334 \text{ kJ/kg}) = 3.34 \times 10^5 \text{ J}.$$

Now the change in entropy is positive, since heat transfer occurs into the ice to cause the phase change; thus,

Equation:

$$\Delta S = \frac{Q}{T} = \frac{3.34 \times 10^5 \text{ J}}{T}.$$

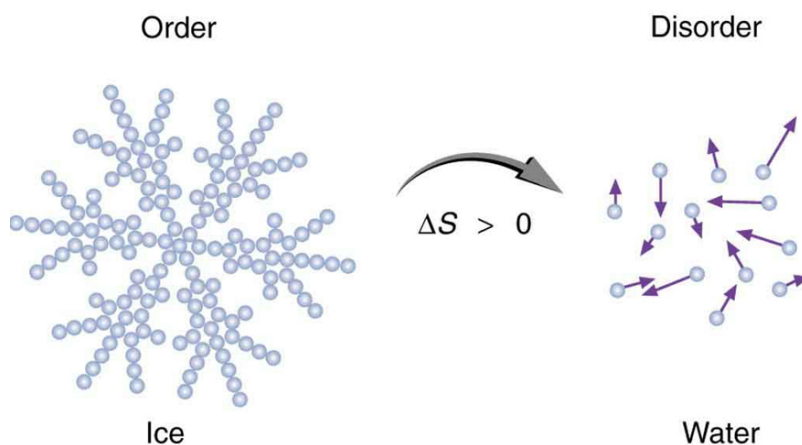
T is the melting temperature of ice. That is, $T = 0^\circ\text{C} = 273 \text{ K}$. So the change in entropy is

Equation:

$$\begin{aligned} \Delta S &= \frac{3.34 \times 10^5 \text{ J}}{273 \text{ K}} \\ &= 1.22 \times 10^3 \text{ J/K}. \end{aligned}$$

Discussion

This is a significant increase in entropy accompanying an increase in disorder.



When ice melts, it becomes more disordered and less structured. The systematic arrangement of molecules in a crystal structure is replaced by a more random and less orderly movement of molecules without

fixed locations or orientations. Its entropy increases because heat transfer occurs into it. Entropy is a measure of disorder.

In another easily imagined example, suppose we mix equal masses of water originally at two different temperatures, say 20.0°C and 40.0°C . The result is water at an intermediate temperature of 30.0°C . Three outcomes have resulted: entropy has increased, some energy has become unavailable to do work, and the system has become less orderly. Let us think about each of these results.

First, entropy has increased for the same reason that it did in the example above. Mixing the two bodies of water has the same effect as heat transfer from the hot one and the same heat transfer into the cold one. The mixing decreases the entropy of the hot water but increases the entropy of the cold water by a greater amount, producing an overall increase in entropy.

Second, once the two masses of water are mixed, there is only one temperature—you cannot run a heat engine with them. The energy that could have been used to run a heat engine is now unavailable to do work.

Third, the mixture is less orderly, or to use another term, less structured. Rather than having two masses at different temperatures and with different distributions of molecular speeds, we now have a single mass with a uniform temperature.

These three results—entropy, unavailability of energy, and disorder—are not only related but are in fact essentially equivalent.

Life, Evolution, and the Second Law of Thermodynamics

Some people misunderstand the second law of thermodynamics, stated in terms of entropy, to say that the process of the evolution of life violates this law. Over time, complex organisms evolved from much simpler ancestors, representing a large decrease in entropy of the Earth's biosphere. It is a fact

that living organisms have evolved to be highly structured, and much lower in entropy than the substances from which they grow. But it is *always* possible for the entropy of one part of the universe to decrease, provided the total change in entropy of the universe increases. In equation form, we can write this as

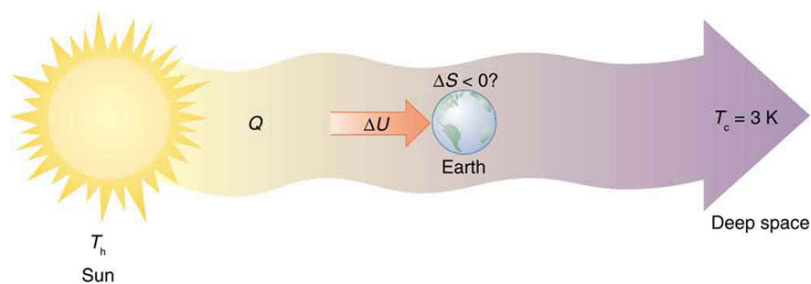
Equation:

$$\Delta S_{\text{tot}} = \Delta S_{\text{syst}} + \Delta S_{\text{envir}} > 0.$$

Thus ΔS_{syst} can be negative as long as ΔS_{envir} is positive and greater in magnitude.

How is it possible for a system to decrease its entropy? Energy transfer is necessary. If I pick up marbles that are scattered about the room and put them into a cup, my work has decreased the entropy of that system. If I gather iron ore from the ground and convert it into steel and build a bridge, my work has decreased the entropy of that system. Energy coming from the Sun can decrease the entropy of local systems on Earth—that is, ΔS_{syst} is negative. But the overall entropy of the rest of the universe increases by a greater amount—that is, ΔS_{envir} is positive and greater in magnitude. Thus, $\Delta S_{\text{tot}} = \Delta S_{\text{syst}} + \Delta S_{\text{envir}} > 0$, and the second law of thermodynamics is *not* violated.

Every time a plant stores some solar energy in the form of chemical potential energy, or an updraft of warm air lifts a soaring bird, the Earth can be viewed as a heat engine operating between a hot reservoir supplied by the Sun and a cold reservoir supplied by dark outer space—a heat engine of high complexity, causing local decreases in entropy as it uses part of the heat transfer from the Sun into deep space. There is a large total increase in entropy resulting from this massive heat transfer. A small part of this heat transfer is stored in structured systems on Earth, producing much smaller local decreases in entropy. (See [\[link\]](#).)



Earth's entropy may decrease in the process of intercepting a small part of the heat transfer from the Sun into deep space.

Entropy for the entire process increases greatly while Earth becomes more structured with living systems and stored energy in various forms.

Note:

PhET Explorations: Reversible Reactions

Watch a reaction proceed over time. How does total energy affect a reaction rate? Vary temperature, barrier height, and potential energies. Record concentrations and time in order to extract rate coefficients. Do temperature dependent studies to extract Arrhenius parameters. This simulation is best used with teacher guidance because it presents an analogy of chemical reactions.

[Reversible Reactions](#)

Section Summary

- Entropy is the loss of energy available to do work.
- Another form of the second law of thermodynamics states that the total entropy of a system either increases or remains constant; it never decreases.
- Entropy is zero in a reversible process; it increases in an irreversible process.
- The ultimate fate of the universe is likely to be thermodynamic equilibrium, where the universal temperature is constant and no energy is available to do work.
- Entropy is also associated with the tendency toward disorder in a closed system.

Conceptual Questions

Exercise:

Problem:

A woman shuts her summer cottage up in September and returns in June. No one has entered the cottage in the meantime. Explain what she is likely to find, in terms of the second law of thermodynamics.

Exercise:

Problem:

Consider a system with a certain energy content, from which we wish to extract as much work as possible. Should the system's entropy be high or low? Is this orderly or disorderly? Structured or uniform? Explain briefly.

Exercise:

Problem:

Does a gas become more orderly when it liquefies? Does its entropy change? If so, does the entropy increase or decrease? Explain your answer.

Exercise:

Problem:

Explain how water's entropy can decrease when it freezes without violating the second law of thermodynamics. Specifically, explain what happens to the entropy of its surroundings.

Exercise:

Problem:

Is a uniform-temperature gas more or less orderly than one with several different temperatures? Which is more structured? In which can heat transfer result in work done without heat transfer from another system?

Exercise:

Problem:

Give an example of a spontaneous process in which a system becomes less ordered and energy becomes less available to do work. What happens to the system's entropy in this process?

Exercise:

Problem:

What is the change in entropy in an adiabatic process? Does this imply that adiabatic processes are reversible? Can a process be precisely adiabatic for a macroscopic system?

Exercise:

Problem:

Does the entropy of a star increase or decrease as it radiates? Does the entropy of the space into which it radiates (which has a temperature of about 3 K) increase or decrease? What does this do to the entropy of the universe?

Exercise:

Problem:

Explain why a building made of bricks has smaller entropy than the same bricks in a disorganized pile. Do this by considering the number of ways that each could be formed (the number of microstates in each macrostate).

Problem Exercises**Exercise:****Problem:**

(a) On a winter day, a certain house loses 5.00×10^8 J of heat to the outside (about 500,000 Btu). What is the total change in entropy due to this heat transfer alone, assuming an average indoor temperature of 21.0°C and an average outdoor temperature of 5.00°C ? (b) This large change in entropy implies a large amount of energy has become unavailable to do work. Where do we find more energy when such energy is lost to us?

Solution:

(a) 9.78×10^4 J/K

(b) In order to gain more energy, we must generate it from things within the house, like a heat pump, human bodies, and other appliances. As you know, we use a lot of energy to keep our houses warm in the winter because of the loss of heat to the outside.

Exercise:**Problem:**

On a hot summer day, 4.00×10^6 J of heat transfer into a parked car takes place, increasing its temperature from 35.0°C to 45.0°C . What is the increase in entropy of the car due to this heat transfer alone?

Exercise:

Problem:

A hot rock ejected from a volcano's lava fountain cools from 1100°C to 40.0°C , and its entropy decreases by 950 J/K . How much heat transfer occurs from the rock?

Solution:

$$8.01 \times 10^5 \text{ J}$$

Exercise:**Problem:**

When $1.60 \times 10^5 \text{ J}$ of heat transfer occurs into a meat pie initially at 20.0°C , its entropy increases by 480 J/K . What is its final temperature?

Exercise:**Problem:**

The Sun radiates energy at the rate of $3.80 \times 10^{26} \text{ W}$ from its 5500°C surface into dark empty space (a negligible fraction radiates onto Earth and the other planets). The effective temperature of deep space is -270°C . (a) What is the increase in entropy in one day due to this heat transfer? (b) How much work is made unavailable?

Solution:

(a) $1.04 \times 10^{31} \text{ J/K}$

(b) $3.28 \times 10^{31} \text{ J}$

Exercise:

Problem:

(a) In reaching equilibrium, how much heat transfer occurs from 1.00 kg of water at 40.0°C when it is placed in contact with 1.00 kg of 20.0°C water in reaching equilibrium? (b) What is the change in entropy due to this heat transfer? (c) How much work is made unavailable, taking the lowest temperature to be 20.0°C ? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Entropy](#).

Exercise:**Problem:**

What is the decrease in entropy of 25.0 g of water that condenses on a bathroom mirror at a temperature of 35.0°C , assuming no change in temperature and given the latent heat of vaporization to be 2450 kJ/kg?

Solution:

199 J/K

Exercise:**Problem:**

Find the increase in entropy of 1.00 kg of liquid nitrogen that starts at its boiling temperature, boils, and warms to 20.0°C at constant pressure.

Exercise:

Problem:

A large electrical power station generates 1000 MW of electricity with an efficiency of 35.0%. (a) Calculate the heat transfer to the power station, Q_h , in one day. (b) How much heat transfer Q_c occurs to the environment in one day? (c) If the heat transfer in the cooling towers is from 35.0°C water into the local air mass, which increases in temperature from 18.0°C to 20.0°C , what is the total increase in entropy due to this heat transfer? (d) How much energy becomes unavailable to do work because of this increase in entropy, assuming an 18.0°C lowest temperature? (Part of Q_c could be utilized to operate heat engines or for simply heating the surroundings, but it rarely is.)

Solution:

(a) $2.47 \times 10^{14}\text{ J}$

(b) $1.60 \times 10^{14}\text{ J}$

(c) $2.85 \times 10^{10}\text{ J/K}$

(d) $8.29 \times 10^{12}\text{ J}$

Exercise:

Problem:

(a) How much heat transfer occurs from 20.0 kg of 90.0° C water placed in contact with 20.0 kg of 10.0° C water, producing a final temperature of 50.0° C? (b) How much work could a Carnot engine do with this heat transfer, assuming it operates between two reservoirs at constant temperatures of 90.0° C and 10.0° C? (c) What increase in entropy is produced by mixing 20.0 kg of 90.0° C water with 20.0 kg of 10.0° C water? (d) Calculate the amount of work made unavailable by this mixing using a low temperature of 10.0° C, and compare it with the work done by the Carnot engine. Explicitly show how you follow the steps in the [Problem-Solving Strategies for Entropy](#). (e) Discuss how everyday processes make increasingly more energy unavailable to do work, as implied by this problem.

Glossary**entropy**

a measurement of a system's disorder and its inability to do work in a system

change in entropy

the ratio of heat transfer to temperature Q/T

second law of thermodynamics stated in terms of entropy

the total entropy of a system either increases or remains constant; it never decreases

Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation

- Identify probabilities in entropy.
- Analyze statistical probabilities in entropic systems.



When you toss a coin a large number of times, heads and tails tend to come up in roughly equal numbers. Why doesn't heads come up 100, 90, or even 80% of the time? (credit: Jon Sullivan, PDPhoto.org)

The various ways of formulating the second law of thermodynamics tell what happens rather than why it happens. Why should heat transfer occur only from hot to cold? Why should energy become ever less available to do work? Why should the universe become increasingly disorderly? The answer is that it is a matter of overwhelming probability. Disorder is simply vastly more likely than order.

When you watch an emerging rain storm begin to wet the ground, you will notice that the drops fall in a disorganized manner both in time and in space. Some fall close together, some far apart, but they never fall in

straight, orderly rows. It is not impossible for rain to fall in an orderly pattern, just highly unlikely, because there are many more disorderly ways than orderly ones. To illustrate this fact, we will examine some random processes, starting with coin tosses.

Coin Tosses

What are the possible outcomes of tossing 5 coins? Each coin can land either heads or tails. On the large scale, we are concerned only with the total heads and tails and not with the order in which heads and tails appear. The following possibilities exist:

Equation:

5 heads, 0 tails

4 heads, 1 tail

3 heads, 2 tails

2 heads, 3 tails

1 head, 4 tails

0 head, 5 tails

These are what we call macrostates. A **macrostate** is an overall property of a system. It does not specify the details of the system, such as the order in which heads and tails occur or which coins are heads or tails.

Using this nomenclature, a system of 5 coins has the 6 possible macrostates just listed. Some macrostates are more likely to occur than others. For instance, there is only one way to get 5 heads, but there are several ways to get 3 heads and 2 tails, making the latter macrostate more probable. [\[link\]](#) lists of all the ways in which 5 coins can be tossed, taking into account the order in which heads and tails occur. Each sequence is called a **microstate**—a detailed description of every element of a system.

	Individual microstates	Number of microstates
5 heads, 0 tails	HHHHH	1
4 heads, 1 tail	HHHHT, HHHTH, HHTHH, HTHHH, THHHH	5
3 heads, 2 tails	HTHTH, THTHH, HTHHT, THHTH, THHHT HTHTH, THTHH, HTHHT, THHTH, THHHT	10
2 heads, 3 tails	TTTHH, TTHHT, THHTT, HHTTT, TTHTH, THTHT, HTHTT, THTTH, HTTHT, HTTTH	10
1 head, 4 tails	TTTTH, TTTHT, TTHTT, THTTT, HTTTT	5
0 heads, 5 tails	TTTTT	1
		Total: 32

5-Coin Toss

The macrostate of 3 heads and 2 tails can be achieved in 10 ways and is thus 10 times more probable than the one having 5 heads. Not surprisingly, it is equally probable to have the reverse, 2 heads and 3 tails. Similarly, it is equally probable to get 5 tails as it is to get 5 heads. Note that all of these conclusions are based on the crucial assumption that each microstate is equally probable. With coin tosses, this requires that the coins not be

asymmetric in a way that favors one side over the other, as with loaded dice. With any system, the assumption that all microstates are equally probable must be valid, or the analysis will be erroneous.

The two most orderly possibilities are 5 heads or 5 tails. (They are more structured than the others.) They are also the least likely, only 2 out of 32 possibilities. The most disorderly possibilities are 3 heads and 2 tails and its reverse. (They are the least structured.) The most disorderly possibilities are also the most likely, with 20 out of 32 possibilities for the 3 heads and 2 tails and its reverse. If we start with an orderly array like 5 heads and toss the coins, it is very likely that we will get a less orderly array as a result, since 30 out of the 32 possibilities are less orderly. So even if you start with an orderly state, there is a strong tendency to go from order to disorder, from low entropy to high entropy. The reverse can happen, but it is unlikely.

Macrostate		Number of microstates
Heads	Tails	(W)
100	0	1
99	1	1.0×10^2
95	5	7.5×10^7
90	10	1.7×10^{13}

Macrostate		Number of microstates
75	25	2.4×10^{23}
60	40	1.4×10^{28}
55	45	6.1×10^{28}
51	49	9.9×10^{28}
50	50	1.0×10^{29}
49	51	9.9×10^{28}
45	55	6.1×10^{28}
40	60	1.4×10^{28}
25	75	2.4×10^{23}

Macrostate		Number of microstates
10	90	1.7×10^{13}
5	95	7.5×10^7
1	99	1.0×10^2
0	100	1
		Total: 1.27×10^{30}

100-Coin Toss

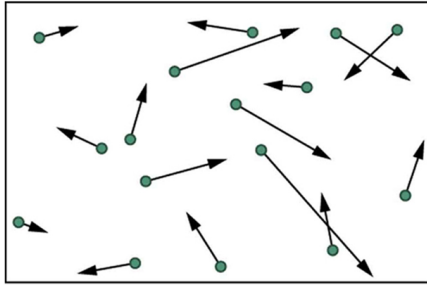
This result becomes dramatic for larger systems. Consider what happens if you have 100 coins instead of just 5. The most orderly arrangements (most structured) are 100 heads or 100 tails. The least orderly (least structured) is that of 50 heads and 50 tails. There is only 1 way (1 microstate) to get the most orderly arrangement of 100 heads. There are 100 ways (100 microstates) to get the next most orderly arrangement of 99 heads and 1 tail (also 100 to get its reverse). And there are 1.0×10^{29} ways to get 50 heads and 50 tails, the least orderly arrangement. [\[link\]](#) is an abbreviated list of the various macrostates and the number of microstates for each macrostate. The total number of microstates—the total number of different ways 100 coins can be tossed—is an impressively large 1.27×10^{30} . Now, if we start with an orderly macrostate like 100 heads and toss the coins, there is a virtual certainty that we will get a less orderly macrostate. If we keep tossing the coins, it is possible, but exceedingly unlikely, that we will ever

get back to the most orderly macrostate. If you tossed the coins once each second, you could expect to get either 100 heads or 100 tails once in 2×10^{22} years! This period is 1 trillion (10^{12}) times longer than the age of the universe, and so the chances are essentially zero. In contrast, there is an 8% chance of getting 50 heads, a 73% chance of getting from 45 to 55 heads, and a 96% chance of getting from 40 to 60 heads. Disorder is highly likely.

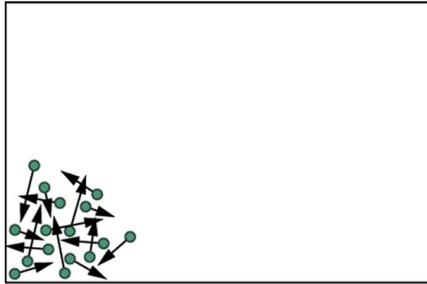
Disorder in a Gas

The fantastic growth in the odds favoring disorder that we see in going from 5 to 100 coins continues as the number of entities in the system increases. Let us now imagine applying this approach to perhaps a small sample of gas. Because counting microstates and macrostates involves statistics, this is called **statistical analysis**. The macrostates of a gas correspond to its macroscopic properties, such as volume, temperature, and pressure; and its microstates correspond to the detailed description of the positions and velocities of its atoms. Even a small amount of gas has a huge number of atoms: 1.0 cm^3 of an ideal gas at 1.0 atm and 0° C has 2.7×10^{19} atoms. So each macrostate has an immense number of microstates. In plain language, this means that there are an immense number of ways in which the atoms in a gas can be arranged, while still having the same pressure, temperature, and so on.

The most likely conditions (or macrostates) for a gas are those we see all the time—a random distribution of atoms in space with a Maxwell-Boltzmann distribution of speeds in random directions, as predicted by kinetic theory. This is the most disorderly and least structured condition we can imagine. In contrast, one type of very orderly and structured macrostate has all of the atoms in one corner of a container with identical velocities. There are very few ways to accomplish this (very few microstates corresponding to it), and so it is exceedingly unlikely ever to occur. (See [\[link\]](#)(b).) Indeed, it is so unlikely that we have a law saying that it is impossible, which has never been observed to be violated—the second law of thermodynamics.



(a) Likely



(b) Highly unlikely

(a) The ordinary state of gas in a container is a disorderly, random distribution of atoms or molecules with a Maxwell-Boltzmann distribution of speeds. It is so unlikely that these atoms or molecules would ever end up in one corner of the container that it might as well be impossible. (b) With energy transfer, the gas can be forced into one corner and its entropy greatly reduced. But left alone, it will

spontaneously
increase its entropy
and return to the
normal conditions,
because they are
immensely more
likely.

The disordered condition is one of high entropy, and the ordered one has low entropy. With a transfer of energy from another system, we could force all of the atoms into one corner and have a local decrease in entropy, but at the cost of an overall increase in entropy of the universe. If the atoms start out in one corner, they will quickly disperse and become uniformly distributed and will never return to the orderly original state ([\[link\]](#)(b)). Entropy will increase. With such a large sample of atoms, it is possible—but unimaginably unlikely—for entropy to decrease. Disorder is vastly more likely than order.

The arguments that disorder and high entropy are the most probable states are quite convincing. The great Austrian physicist Ludwig Boltzmann (1844–1906)—who, along with Maxwell, made so many contributions to kinetic theory—proved that the entropy of a system in a given state (a macrostate) can be written as

Equation:

$$S = k \ln W,$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate. W is proportional to the probability that the macrostate will occur. Thus entropy is directly related to the probability of a state—the more likely the state, the greater its entropy. Boltzmann proved that this expression for S is equivalent to the definition $\Delta S = Q/T$, which we have used extensively.

Thus the second law of thermodynamics is explained on a very basic level: entropy either remains the same or increases in every process. This phenomenon is due to the extraordinarily small probability of a decrease, based on the extraordinarily larger number of microstates in systems with greater entropy. Entropy *can* decrease, but for any macroscopic system, this outcome is so unlikely that it will never be observed.

Example:**Entropy Increases in a Coin Toss**

Suppose you toss 100 coins starting with 60 heads and 40 tails, and you get the most likely result, 50 heads and 50 tails. What is the change in entropy?

Strategy

Noting that the number of microstates is labeled W in [\[link\]](#) for the 100-coin toss, we can use $\Delta S = S_f - S_i = k \ln W_f - k \ln W_i$ to calculate the change in entropy.

Solution

The change in entropy is

Equation:

$$\Delta S = S_f - S_i = k \ln W_f - k \ln W_i,$$

where the subscript i stands for the initial 60 heads and 40 tails state, and the subscript f for the final 50 heads and 50 tails state. Substituting the values for W from [\[link\]](#) gives

Equation:

$$\begin{aligned}\Delta S &= (1.38 \times 10^{-23} \text{ J/K}) [\ln(1.0 \times 10^{29}) - \ln(1.4 \times 10^{28})] \\ &= 2.7 \times 10^{-23} \text{ J/K}\end{aligned}$$

Discussion

This increase in entropy means we have moved to a less orderly situation. It is not impossible for further tosses to produce the initial state of 60 heads and 40 tails, but it is less likely. There is about a 1 in 90 chance for that decrease in entropy ($-2.7 \times 10^{-23} \text{ J/K}$) to occur. If we calculate the decrease in entropy to move to the most orderly state, we get

$\Delta S = -92 \times 10^{-23} \text{ J/K}$. There is about a 1 in 10^{30} chance of this change occurring. So while very small decreases in entropy are unlikely, slightly greater decreases are impossibly unlikely. These probabilities imply, again, that for a macroscopic system, a decrease in entropy is impossible. For example, for heat transfer to occur spontaneously from 1.00 kg of 0°C ice to its 0°C environment, there would be a decrease in entropy of $1.22 \times 10^3 \text{ J/K}$. Given that a ΔS of 10^{-21} J/K corresponds to about a 1 in 10^{30} chance, a decrease of this size (10^3 J/K) is an *utter* impossibility. Even for a milligram of melted ice to spontaneously refreeze is impossible.

Note:

Problem-Solving Strategies for Entropy

1. *Examine the situation to determine if entropy is involved.*
2. *Identify the system of interest and draw a labeled diagram of the system showing energy flow.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful.*
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns). You must carefully identify the heat transfer, if any, and the temperature at which the process takes place. It is also important to identify the initial and final states.*
5. *Solve the appropriate equation for the quantity to be determined (the unknown). Note that the change in entropy can be determined between any states by calculating it for a reversible process.*
6. *Substitute the known value along with their units into the appropriate equation, and obtain numerical solutions complete with units.*
7. *To see if it is reasonable: Does it make sense?* For example, total entropy should increase for any real process or be constant for a reversible process. Disordered states should be more probable and have greater entropy than ordered states.

Section Summary

- Disorder is far more likely than order, which can be seen statistically.
- The entropy of a system in a given state (a macrostate) can be written as

Equation:

$$S = k \ln W,$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate.

Conceptual Questions

Exercise:

Problem:

Explain why a building made of bricks has smaller entropy than the same bricks in a disorganized pile. Do this by considering the number of ways that each could be formed (the number of microstates in each macrostate).

Problem Exercises

Exercise:

Problem:

Using [\[link\]](#), verify the contention that if you toss 100 coins each second, you can expect to get 100 heads or 100 tails once in 2×10^{22} years; calculate the time to two-digit accuracy.

Solution:

It should happen twice in every 1.27×10^{30} s or once in every

$$6.35 \times 10^{29} \text{ s} \left(6.35 \times 10^{29} \text{ s} \right) \left(\frac{1 \text{ h}}{3600 \text{ s}} \right) \left(\frac{1 \text{ d}}{24 \text{ h}} \right) \left(\frac{1 \text{ y}}{365.25 \text{ d}} \right)$$

$$= 2.0 \times 10^{22} \text{ y}$$

Exercise:

Problem:

What percent of the time will you get something in the range from 60 heads and 40 tails through 40 heads and 60 tails when tossing 100 coins? The total number of microstates in that range is 1.22×10^{30} . (Consult [\[link\]](#).)

Exercise:

Problem:

(a) If tossing 100 coins, how many ways (microstates) are there to get the three most likely macrostates of 49 heads and 51 tails, 50 heads and 50 tails, and 51 heads and 49 tails? (b) What percent of the total possibilities is this? (Consult [\[link\]](#).)

Solution:

(a) 3.0×10^{29}

(b) 24%

Exercise:

Problem:

(a) What is the change in entropy if you start with 100 coins in the 45 heads and 55 tails macrostate, toss them, and get 51 heads and 49 tails? (b) What if you get 75 heads and 25 tails? (c) How much more likely is 51 heads and 49 tails than 75 heads and 25 tails? (d) Does either outcome violate the second law of thermodynamics?

Exercise:

Problem:

(a) What is the change in entropy if you start with 10 coins in the 5 heads and 5 tails macrostate, toss them, and get 2 heads and 8 tails? (b) How much more likely is 5 heads and 5 tails than 2 heads and 8 tails? (Take the ratio of the number of microstates to find out.) (c) If you were betting on 2 heads and 8 tails would you accept odds of 252 to 45? Explain why or why not.

Solution:

(a) $-2.38 \times 10^{-23} \text{ J/K}$

(b) 5.6 times more likely

(c) If you were betting on two heads and 8 tails, the odds of breaking even are 252 to 45, so on average you would break even. So, no, you wouldn't bet on odds of 252 to 45.

Macrostate		Number of Microstates (W)
Heads	Tails	
10	0	1
9	1	10
8	2	45
7	3	120

Macrostate		Number of Microstates (W)
6	4	210
5	5	252
4	6	210
3	7	120
2	8	45
1	9	10
0	10	1
		Total: 1024

10-Coin Toss

Exercise:

Problem:

(a) If you toss 10 coins, what percent of the time will you get the three most likely macrostates (6 heads and 4 tails, 5 heads and 5 tails, 4 heads and 6 tails)? (b) You can realistically toss 10 coins and count the number of heads and tails about twice a minute. At that rate, how long will it take on average to get either 10 heads and 0 tails or 0 heads and 10 tails?

Exercise:

Problem:

(a) Construct a table showing the macrostates and all of the individual microstates for tossing 6 coins. (Use [\[link\]](#) as a guide.) (b) How many macrostates are there? (c) What is the total number of microstates? (d) What percent chance is there of tossing 5 heads and 1 tail? (e) How much more likely are you to toss 3 heads and 3 tails than 5 heads and 1 tail? (Take the ratio of the number of microstates to find out.)

Solution:

(b) 7

(c) 64

(d) 9.38%

(e) 3.33 times more likely (20 to 6)

Exercise:**Problem:**

In an air conditioner, 12.65 MJ of heat transfer occurs from a cold environment in 1.00 h. (a) What mass of ice melting would involve the same heat transfer? (b) How many hours of operation would be equivalent to melting 900 kg of ice? (c) If ice costs 20 cents per kg, do you think the air conditioner could be operated more cheaply than by simply using ice? Describe in detail how you evaluate the relative costs.

Glossary

macrostate

an overall property of a system

microstate

each sequence within a larger macrostate

statistical analysis

using statistics to examine data, such as counting microstates and macrostates

Introduction to Oscillatory Motion and Waves

class="introduction"

There
are at
least
four
types
of
waves
in this
picture
—only
the
water
waves
are
evident
. There
are also
sound
waves,
light
waves,
and
waves
on the
guitar
strings.
(credit:
John
Norton
)



What do an ocean buoy, a child in a swing, the cone inside a speaker, a guitar, atoms in a crystal, the motion of chest cavities, and the beating of hearts all have in common? They all **oscillate**—that is, they move back and forth between two points. Many systems oscillate, and they have certain characteristics in common. All oscillations involve force and energy. You push a child in a swing to get the motion started. The energy of atoms vibrating in a crystal can be increased with heat. You put energy into a guitar string when you pluck it.

Some oscillations create **waves**. A guitar creates sound waves. You can make water waves in a swimming pool by slapping the water with your hand. You can no doubt think of other types of waves. Some, such as water waves, are visible. Some, such as sound waves, are not. But *every wave is a disturbance that moves from its source and carries energy*. Other examples of waves include earthquakes and visible light. Even subatomic particles, such as electrons, can behave like waves.

By studying oscillatory motion and waves, we shall find that a small number of underlying principles describe all of them and that wave phenomena are more common than you have ever imagined. We begin by studying the type of force that underlies the simplest oscillations and waves. We will then expand our exploration of oscillatory motion and waves to

include concepts such as simple harmonic motion, uniform circular motion, and damped harmonic motion. Finally, we will explore what happens when two or more waves share the same space, in the phenomena known as superposition and interference.

Glossary

oscillate

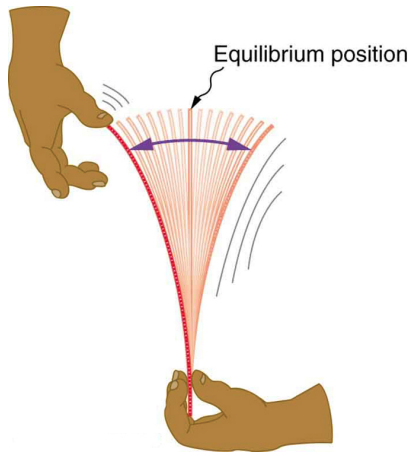
moving back and forth regularly between two points

wave

a disturbance that moves from its source and carries energy

Hooke's Law: Stress and Strain Revisited

- Explain Newton's third law of motion with respect to stress and deformation.
- Describe the restoration of force and displacement.
- Calculate the energy in Hooke's Law of deformation, and the stored energy in a spring.



When displaced from its vertical equilibrium position, this plastic ruler oscillates back and forth because of the restoring force opposing displacement.

When the ruler is on the left, there is a force to the right, and vice versa.

Newton's first law implies that an object oscillating back and forth is experiencing forces. Without force, the object would move in a straight line

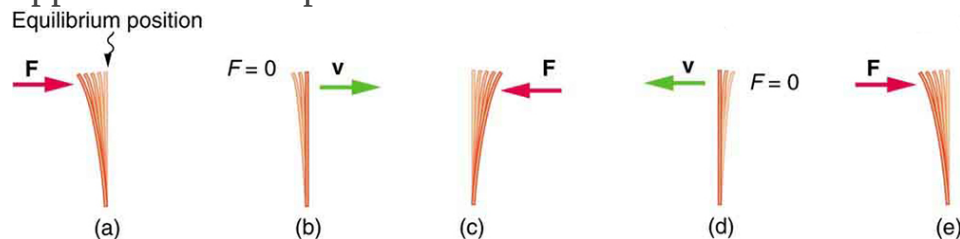
at a constant speed rather than oscillate. Consider, for example, plucking a plastic ruler to the left as shown in [\[link\]](#). The deformation of the ruler creates a force in the opposite direction, known as a **restoring force**. Once released, the restoring force causes the ruler to move back toward its stable equilibrium position, where the net force on it is zero. However, by the time the ruler gets there, it gains momentum and continues to move to the right, producing the opposite deformation. It is then forced to the left, back through equilibrium, and the process is repeated until dissipative forces dampen the motion. These forces remove mechanical energy from the system, gradually reducing the motion until the ruler comes to rest.

The simplest oscillations occur when the restoring force is directly proportional to displacement. When stress and strain were covered in [Newton's Third Law of Motion](#), the name was given to this relationship between force and displacement was Hooke's law:

Equation:

$$F = -kx.$$

Here, F is the restoring force, x is the displacement from equilibrium or **deformation**, and k is a constant related to the difficulty in deforming the system. The minus sign indicates the restoring force is in the direction opposite to the displacement.

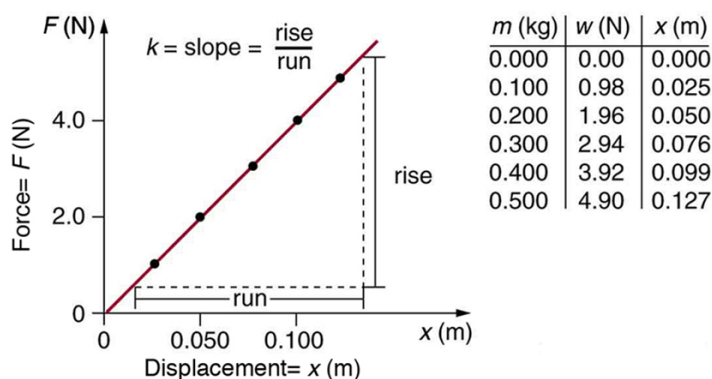


(a) The plastic ruler has been released, and the restoring force is returning the ruler to its equilibrium position. (b) The net force is zero at the equilibrium position, but the ruler has momentum and continues to move to the right. (c) The restoring force is in the opposite direction. It stops the ruler and moves it back toward equilibrium again. (d) Now the ruler has momentum to the left. (e) In the absence of damping

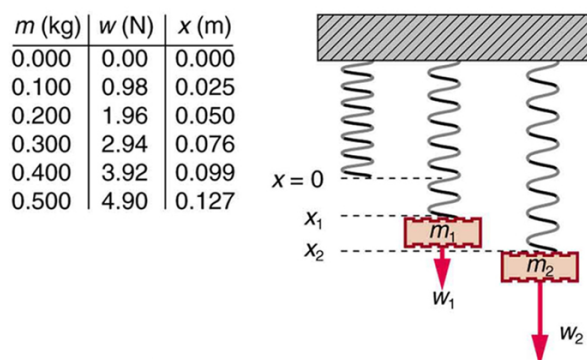
(caused by frictional forces), the ruler reaches its original position. From there, the motion will repeat itself.

The **force constant** k is related to the rigidity (or stiffness) of a system—the larger the force constant, the greater the restoring force, and the stiffer the system. The units of k are newtons per meter (N/m). For example, k is directly related to Young’s modulus when we stretch a string. [\[link\]](#) shows a graph of the absolute value of the restoring force versus the displacement for a system that can be described by Hooke’s law—a simple spring in this case. The slope of the graph equals the force constant k in newtons per meter. A common physics laboratory exercise is to measure restoring forces created by springs, determine if they follow Hooke’s law, and calculate their force constants if they do.

a)



b)



(a) A graph of absolute value of the restoring force versus displacement is

displayed. The fact that the graph is a straight line means that the system obeys Hooke's law. The slope of the graph is the force constant k . (b) The data in the graph were generated by measuring the displacement of a spring from equilibrium while supporting various weights. The restoring force equals the weight supported, if the mass is stationary.

Example:
How Stiff Are Car Springs?



The mass of a car increases due to the introduction of a passenger. This affects the displacement of

the car on its
suspension
system. (credit:
exfordy on
Flickr)

What is the force constant for the suspension system of a car that settles 1.20 cm when an 80.0-kg person gets in?

Strategy

Consider the car to be in its equilibrium position $x = 0$ before the person gets in. The car then settles down 1.20 cm, which means it is displaced to a position $x = -1.20 \times 10^{-2}$ m. At that point, the springs supply a restoring force F equal to the person's weight

$w = mg = (80.0 \text{ kg})(9.80 \text{ m/s}^2) = 784 \text{ N}$. We take this force to be F in Hooke's law. Knowing F and x , we can then solve the force constant k .

Solution

1. Solve Hooke's law, $F = -kx$, for k :

Equation:

$$k = -\frac{F}{x}.$$

Substitute known values and solve k :

Equation:

$$\begin{aligned} k &= -\frac{784 \text{ N}}{-1.20 \times 10^{-2} \text{ m}} \\ &= 6.53 \times 10^4 \text{ N/m.} \end{aligned}$$

Discussion

Note that F and x have opposite signs because they are in opposite directions—the restoring force is up, and the displacement is down. Also, note that the car would oscillate up and down when the person got in if it

were not for damping (due to frictional forces) provided by shock absorbers. Bouncing cars are a sure sign of bad shock absorbers.

Energy in Hooke's Law of Deformation

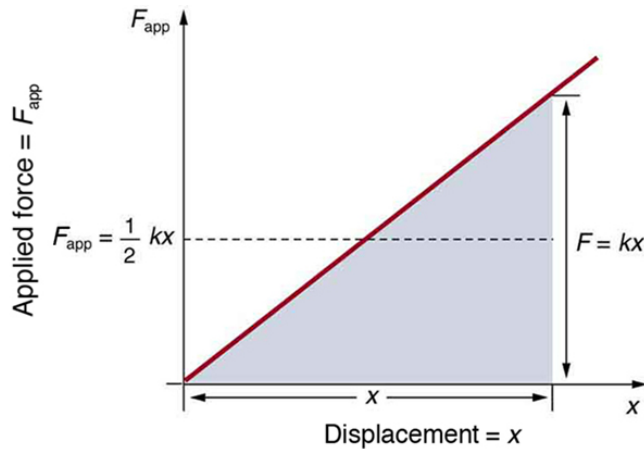
In order to produce a deformation, work must be done. That is, a force must be exerted through a distance, whether you pluck a guitar string or compress a car spring. If the only result is deformation, and no work goes into thermal, sound, or kinetic energy, then all the work is initially stored in the deformed object as some form of potential energy. The potential energy stored in a spring is $PE_{\text{el}} = \frac{1}{2}kx^2$. Here, we generalize the idea to elastic potential energy for a deformation of any system that can be described by Hooke's law. Hence,

Equation:

$$PE_{\text{el}} = \frac{1}{2}kx^2,$$

where PE_{el} is the **elastic potential energy** stored in any deformed system that obeys Hooke's law and has a displacement x from equilibrium and a force constant k .

It is possible to find the work done in deforming a system in order to find the energy stored. This work is performed by an applied force F_{app} . The applied force is exactly opposite to the restoring force (action-reaction), and so $F_{\text{app}} = kx$. [\[link\]](#) shows a graph of the applied force versus deformation x for a system that can be described by Hooke's law. Work done on the system is force multiplied by distance, which equals the area under the curve or $(1/2)kx^2$ (Method A in the figure). Another way to determine the work is to note that the force increases linearly from 0 to kx , so that the average force is $(1/2)kx$, the distance moved is x , and thus $W = F_{\text{app}}d = [(1/2)kx](x) = (1/2)kx^2$ (Method B in the figure).



Method A

$$W = \frac{1}{2} bh = \frac{1}{2} kxx$$

$$W = \frac{1}{2} kx^2$$

Method B

$$W = f \cdot x = \left(\frac{1}{2} kx \right) (x)$$

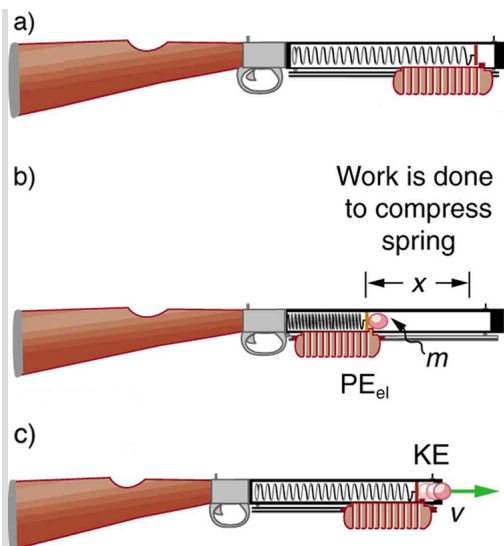
$$W = \frac{1}{2} kx^2$$

A graph of applied force versus distance for the deformation of a system that can be described by Hooke's law is displayed. The work done on the system equals the area under the graph or the area of the triangle, which is half its base multiplied by its height, or $W = (1/2)kx^2$.

Example:

Calculating Stored Energy: A Tranquilizer Gun Spring

We can use a toy gun's spring mechanism to ask and answer two simple questions: (a) How much energy is stored in the spring of a tranquilizer gun that has a force constant of 50.0 N/m and is compressed 0.150 m? (b) If you neglect friction and the mass of the spring, at what speed will a 2.00-g projectile be ejected from the gun?



(a) In this image of the gun, the spring is uncompressed before being cocked. (b) The spring has been compressed a distance x , and the projectile is in place. (c) When released, the spring converts elastic potential energy PE_{el} into kinetic energy.

Strategy for a

(a): The energy stored in the spring can be found directly from elastic potential energy equation, because k and x are given.

Solution for a

Entering the given values for k and x yields

Equation:

$$\begin{aligned} PE_{el} &= \frac{1}{2} kx^2 = \frac{1}{2} (50.0 \text{ N/m})(0.150 \text{ m})^2 = 0.563 \text{ N} \cdot \text{m} \\ &= 0.563 \text{ J} \end{aligned}$$

Strategy for b

Because there is no friction, the potential energy is converted entirely into kinetic energy. The expression for kinetic energy can be solved for the projectile's speed.

Solution for b

1. Identify known quantities:

Equation:

$$KE_f = PE_{el} \text{ or } \frac{1}{2}mv^2 = (1/2)kx^2 = PE_{el} = 0.563 \text{ J}$$

2. Solve for v :

Equation:

$$v = \left[\frac{2PE_{el}}{m} \right]^{1/2} = \left[\frac{2(0.563 \text{ J})}{0.002 \text{ kg}} \right]^{1/2} = 23.7(\text{J/kg})^{1/2}$$

3. Convert units: 23.7 m/s

Discussion

(a) and (b): This projectile speed is impressive for a tranquilizer gun (more than 80 km/h). The numbers in this problem seem reasonable. The force needed to compress the spring is small enough for an adult to manage, and the energy imparted to the dart is small enough to limit the damage it might do. Yet, the speed of the dart is great enough for it to travel an acceptable distance.

Exercise:

Check your Understanding

Problem:

Envision holding the end of a ruler with one hand and deforming it with the other. When you let go, you can see the oscillations of the ruler. In what way could you modify this simple experiment to increase the rigidity of the system?

Solution:

Answer

You could hold the ruler at its midpoint so that the part of the ruler that oscillates is half as long as in the original experiment.

Exercise:

Check your Understanding

Problem:

If you apply a deforming force on an object and let it come to equilibrium, what happened to the work you did on the system?

Solution:

Answer

It was stored in the object as potential energy.

Section Summary

- An oscillation is a back and forth motion of an object between two points of deformation.
- An oscillation may create a wave, which is a disturbance that propagates from where it was created.
- The simplest type of oscillations and waves are related to systems that can be described by Hooke's law:

Equation:

$$F = -kx,$$

where F is the restoring force, x is the displacement from equilibrium or deformation, and k is the force constant of the system.

- Elastic potential energy PE_{el} stored in the deformation of a system that can be described by Hooke's law is given by

Equation:

$$PE_{\text{el}} = (1/2)kx^2.$$

Conceptual Questions

Exercise:

Problem:

Describe a system in which elastic potential energy is stored.

Problems & Exercises

Exercise:

Problem:

Fish are hung on a spring scale to determine their mass (most fishermen feel no obligation to truthfully report the mass).

- (a) What is the force constant of the spring in such a scale if it the spring stretches 8.00 cm for a 10.0 kg load?
- (b) What is the mass of a fish that stretches the spring 5.50 cm?
- (c) How far apart are the half-kilogram marks on the scale?

Solution:

- (a) $1.23 \times 10^3 \text{ N/m}$
- (b) 6.88 kg
- (c) 4.00 mm

Exercise:

Problem:

It is weigh-in time for the local under-85-kg rugby team. The bathroom scale used to assess eligibility can be described by Hooke's law and is depressed 0.75 cm by its maximum load of 120 kg. (a) What is the spring's effective spring constant? (b) A player stands on the scales and depresses it by 0.48 cm. Is he eligible to play on this under-85 kg team?

Exercise:**Problem:**

One type of BB gun uses a spring-driven plunger to blow the BB from its barrel. (a) Calculate the force constant of its plunger's spring if you must compress it 0.150 m to drive the 0.0500-kg plunger to a top speed of 20.0 m/s. (b) What force must be exerted to compress the spring?

Solution:

(a) 889 N/m

(b) 133 N

Exercise:**Problem:**

(a) The springs of a pickup truck act like a single spring with a force constant of 1.30×10^5 N/m. By how much will the truck be depressed by its maximum load of 1000 kg?

(b) If the pickup truck has four identical springs, what is the force constant of each?

Exercise:**Problem:**

When an 80.0-kg man stands on a pogo stick, the spring is compressed 0.120 m.

(a) What is the force constant of the spring? (b) Will the spring be compressed more when he hops down the road?

Solution:

(a) $6.53 \times 10^3 \text{ N/m}$

(b) Yes

Exercise:

Problem:

A spring has a length of 0.200 m when a 0.300-kg mass hangs from it, and a length of 0.750 m when a 1.95-kg mass hangs from it. (a) What is the force constant of the spring? (b) What is the unloaded length of the spring?

Glossary

deformation

displacement from equilibrium

elastic potential energy

potential energy stored as a result of deformation of an elastic object, such as the stretching of a spring

force constant

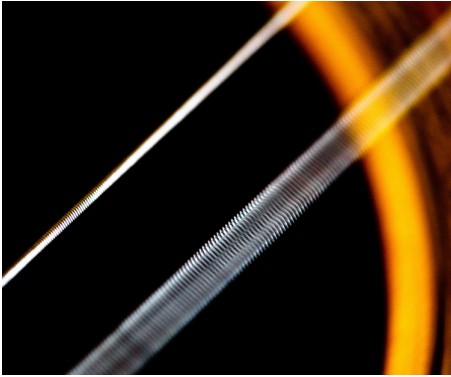
a constant related to the rigidity of a system: the larger the force constant, the more rigid the system; the force constant is represented by k

restoring force

force acting in opposition to the force caused by a deformation

Period and Frequency in Oscillations

- Observe the vibrations of a guitar string.
- Determine the frequency of oscillations.



The strings on this
guitar vibrate at
regular time intervals.
(credit: JAR)

When you pluck a guitar string, the resulting sound has a steady tone and lasts a long time. Each successive vibration of the string takes the same time as the previous one. We define **periodic motion** to be a motion that repeats itself at regular time intervals, such as exhibited by the guitar string or by an object on a spring moving up and down. The time to complete one oscillation remains constant and is called the **period T** . Its units are usually seconds, but may be any convenient unit of time. The word period refers to the time for some event whether repetitive or not; but we shall be primarily interested in periodic motion, which is by definition repetitive. A concept closely related to period is the frequency of an event. For example, if you get a paycheck twice a month, the frequency of payment is two per month and the period between checks is half a month. **Frequency f** is defined to be the number of events per unit time. For periodic motion, frequency is the number of oscillations per unit time. The relationship between frequency and period is

Equation:

$$f = \frac{1}{T}.$$

The SI unit for frequency is the *cycle per second*, which is defined to be a *hertz* (Hz):

Equation:

$$1 \text{ Hz} = 1 \frac{\text{cycle}}{\text{sec}} \text{ or } 1 \text{ Hz} = \frac{1}{\text{s}}$$

A cycle is one complete oscillation. Note that a vibration can be a single or multiple event, whereas oscillations are usually repetitive for a significant number of cycles.

Example:

Determine the Frequency of Two Oscillations: Medical Ultrasound and the Period of Middle C

We can use the formulas presented in this module to determine both the frequency based on known oscillations and the oscillation based on a known frequency. Let's try one example of each. (a) A medical imaging device produces ultrasound by oscillating with a period of $0.400 \mu\text{s}$. What is the frequency of this oscillation? (b) The frequency of middle C on a typical musical instrument is 264 Hz. What is the time for one complete oscillation?

Strategy

Both questions (a) and (b) can be answered using the relationship between period and frequency. In question (a), the period T is given and we are asked to find frequency f . In question (b), the frequency f is given and we are asked to find the period T .

Solution a

1. Substitute $0.400 \mu\text{s}$ for T in $f = \frac{1}{T}$:

Equation:

$$f = \frac{1}{T} = \frac{1}{0.400 \times 10^{-6} \text{ s}}.$$

Solve to find

Equation:

$$f = 2.50 \times 10^6 \text{ Hz}.$$

Discussion a

The frequency of sound found in (a) is much higher than the highest frequency that humans can hear and, therefore, is called ultrasound. Appropriate oscillations at this frequency generate ultrasound used for noninvasive medical diagnoses, such as observations of a fetus in the womb.

Solution b

1. Identify the known values:

The time for one complete oscillation is the period T :

Equation:

$$f = \frac{1}{T}.$$

2. Solve for T :

Equation:

$$T = \frac{1}{f}.$$

3. Substitute the given value for the frequency into the resulting expression:

Equation:

$$T = \frac{1}{f} = \frac{1}{264 \text{ Hz}} = \frac{1}{264 \text{ cycles/s}} = 3.79 \times 10^{-3} \text{ s} = 3.79 \text{ ms}.$$

Discussion

The period found in (b) is the time per cycle, but this value is often quoted as simply the time in convenient units (ms or milliseconds in this case).

Exercise:

Check your Understanding

Problem:

Identify an event in your life (such as receiving a paycheck) that occurs regularly. Identify both the period and frequency of this event.

Solution:

I visit my parents for dinner every other Sunday. The frequency of my visits is 26 per calendar year. The period is two weeks.

Section Summary

- Periodic motion is a repetitious oscillation.
- The time for one oscillation is the period T .
- The number of oscillations per unit time is the frequency f .
- These quantities are related by

Equation:

$$f = \frac{1}{T}.$$

Problems & Exercises

Exercise:

Problem: What is the period of 60.0 Hz electrical power?

Solution:

16.7 ms

Exercise:**Problem:**

If your heart rate is 150 beats per minute during strenuous exercise, what is the time per beat in units of seconds?

Solution:

0.400 s/beats

Exercise:**Problem:**

Find the frequency of a tuning fork that takes 2.50×10^{-3} s to complete one oscillation.

Solution:

400 Hz

Exercise:**Problem:**

A stroboscope is set to flash every 8.00×10^{-5} s. What is the frequency of the flashes?

Solution:

12,500 Hz

Exercise:

Problem:

A tire has a tread pattern with a crevice every 2.00 cm. Each crevice makes a single vibration as the tire moves. What is the frequency of these vibrations if the car moves at 30.0 m/s?

Solution:

1.50 kHz

Exercise:**Problem: Engineering Application**

Each piston of an engine makes a sharp sound every other revolution of the engine. (a) How fast is a race car going if its eight-cylinder engine emits a sound of frequency 750 Hz, given that the engine makes 2000 revolutions per kilometer? (b) At how many revolutions per minute is the engine rotating?

Solution:

(a) 93.8 m/s

(b) 11.3×10^3 rev/min

Glossary

period

time it takes to complete one oscillation

periodic motion

motion that repeats itself at regular time intervals

frequency

number of events per unit of time

Simple Harmonic Motion: A Special Periodic Motion

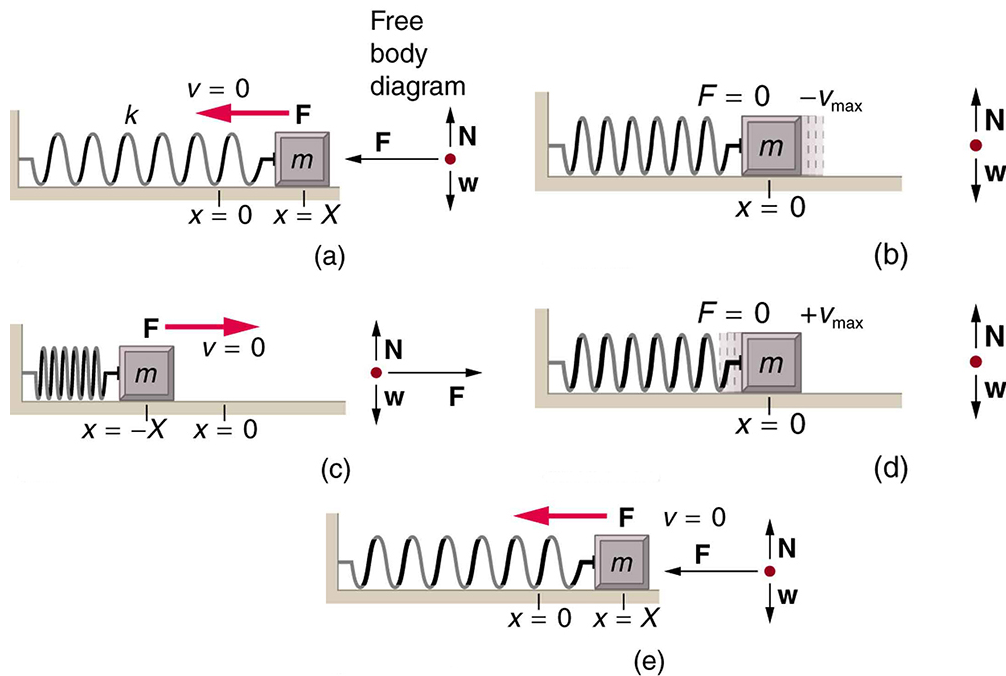
- Describe a simple harmonic oscillator.
- Explain the link between simple harmonic motion and waves.

The oscillations of a system in which the net force can be described by Hooke's law are of special importance, because they are very common. They are also the simplest oscillatory systems. **Simple Harmonic Motion** (SHM) is the name given to oscillatory motion for a system where the net force can be described by Hooke's law, and such a system is called a **simple harmonic oscillator**. If the net force can be described by Hooke's law and there is no *damping* (by friction or other non-conservative forces), then a simple harmonic oscillator will oscillate with equal displacement on either side of the equilibrium position, as shown for an object on a spring in [\[link\]](#). The maximum displacement from equilibrium is called the **amplitude** X . The units for amplitude and displacement are the same, but depend on the type of oscillation. For the object on the spring, the units of amplitude and displacement are meters; whereas for sound oscillations, they have units of pressure (and other types of oscillations have yet other units). Because amplitude is the maximum displacement, it is related to the energy in the oscillation.

Note:

Take-Home Experiment: SHM and the Marble

Find a bowl or basin that is shaped like a hemisphere on the inside. Place a marble inside the bowl and tilt the bowl periodically so the marble rolls from the bottom of the bowl to equally high points on the sides of the bowl. Get a feel for the force required to maintain this periodic motion. What is the restoring force and what role does the force you apply play in the simple harmonic motion (SHM) of the marble?



An object attached to a spring sliding on a frictionless surface is an uncomplicated simple harmonic oscillator. When displaced from equilibrium, the object performs simple harmonic motion that has an amplitude X and a period T . The object's maximum speed occurs as it passes through equilibrium. The stiffer the spring is, the smaller the period T . The greater the mass of the object is, the greater the period T .

What is so significant about simple harmonic motion? One special thing is that the period T and frequency f of a simple harmonic oscillator are independent of amplitude. The string of a guitar, for example, will oscillate with the same frequency whether plucked gently or hard. Because the period is constant, a simple harmonic oscillator can be used as a clock.

Two important factors do affect the period of a simple harmonic oscillator. The period is related to how stiff the system is. A very stiff object has a large force constant k , which causes the system to have a smaller period. For example, you can adjust a diving board's stiffness—the stiffer it is, the

faster it vibrates, and the shorter its period. Period also depends on the mass of the oscillating system. The more massive the system is, the longer the period. For example, a heavy person on a diving board bounces up and down more slowly than a light one.

In fact, the mass m and the force constant k are the *only* factors that affect the period and frequency of simple harmonic motion.

Note:

Period of Simple Harmonic Oscillator

The *period of a simple harmonic oscillator* is given by

Equation:

$$T = 2\pi\sqrt{\frac{m}{k}}$$

and, because $f = 1/T$, the *frequency of a simple harmonic oscillator* is

Equation:

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}.$$

Note that neither T nor f has any dependence on amplitude.

Note:

Take-Home Experiment: Mass and Ruler Oscillations

Find two identical wooden or plastic rulers. Tape one end of each ruler firmly to the edge of a table so that the length of each ruler that protrudes from the table is the same. On the free end of one ruler tape a heavy object such as a few large coins. Pluck the ends of the rulers at the same time and observe which one undergoes more cycles in a time period, and measure the period of oscillation of each of the rulers.

Example:**Calculate the Frequency and Period of Oscillations: Bad Shock Absorbers in a Car**

If the shock absorbers in a car go bad, then the car will oscillate at the least provocation, such as when going over bumps in the road and after stopping (See [\[link\]](#)). Calculate the frequency and period of these oscillations for such a car if the car's mass (including its load) is 900 kg and the force constant (k) of the suspension system is 6.53×10^4 N/m.

Strategy

The frequency of the car's oscillations will be that of a simple harmonic oscillator as given in the equation $f = \frac{1}{2\pi} \sqrt{\frac{k}{m}}$. The mass and the force constant are both given.

Solution

1. Enter the known values of k and m :

Equation:

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{m}} = \frac{1}{2\pi} \sqrt{\frac{6.53 \times 10^4 \text{ N/m}}{900 \text{ kg}}}.$$

2. Calculate the frequency:

Equation:

$$\frac{1}{2\pi} \sqrt{72.6/\text{s}^{-2}} = 1.3656/\text{s}^{-1} \approx 1.36/\text{s}^{-1} = 1.36 \text{ Hz}.$$

3. You could use $T = 2\pi\sqrt{\frac{m}{k}}$ to calculate the period, but it is simpler to use the relationship $T = 1/f$ and substitute the value just found for f :

Equation:

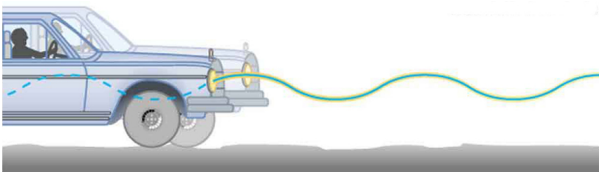
$$T = \frac{1}{f} = \frac{1}{1.356 \text{ Hz}} = 0.738 \text{ s}.$$

Discussion

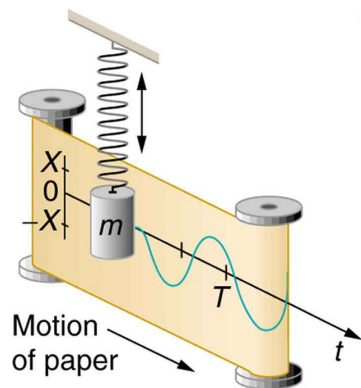
The values of T and f both seem about right for a bouncing car. You can observe these oscillations if you push down hard on the end of a car and let go.

The Link between Simple Harmonic Motion and Waves

If a time-exposure photograph of the bouncing car were taken as it drove by, the headlight would make a wavelike streak, as shown in [\[link\]](#). Similarly, [\[link\]](#) shows an object bouncing on a spring as it leaves a wavelike "trace of its position on a moving strip of paper. Both waves are sine functions. All simple harmonic motion is intimately related to sine and cosine waves.



The bouncing car makes a wavelike motion. If the restoring force in the suspension system can be described only by Hooke's law, then the wave is a sine function. (The wave is the trace produced by the headlight as the car moves to the right.)



The vertical position of an object bouncing on a spring is recorded on a strip of moving paper, leaving a sine wave.

The displacement as a function of time t in any simple harmonic motion—that is, one in which the net restoring force can be described by Hooke's law, is given by

Equation:

$$x(t) = X \cos \frac{2\pi t}{T},$$

where X is amplitude. At $t = 0$, the initial position is $x_0 = X$, and the displacement oscillates back and forth with a period T . (When $t = T$, we get $x = X$ again because $\cos 2\pi = 1$). Furthermore, from this expression for x , the velocity v as a function of time is given by:

Equation:

$$v(t) = -v_{\max} \sin \left(\frac{2\pi t}{T} \right),$$

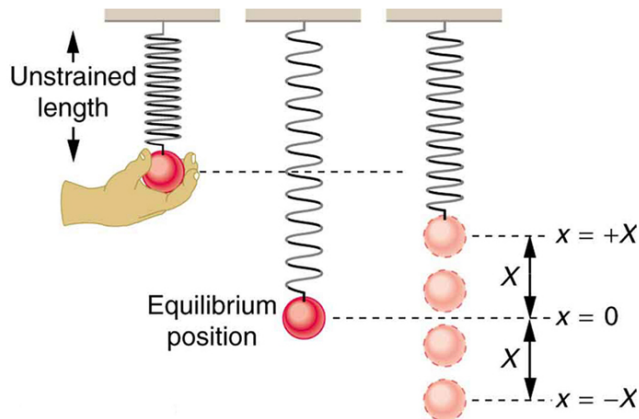
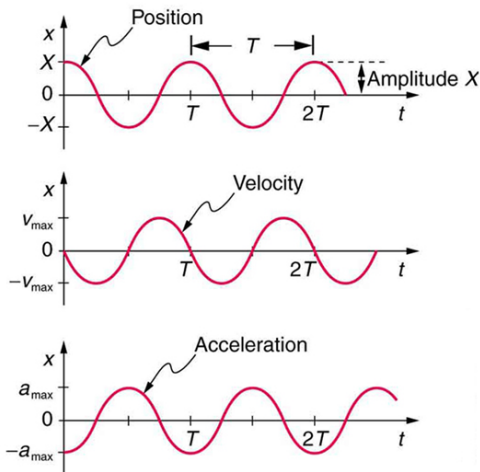
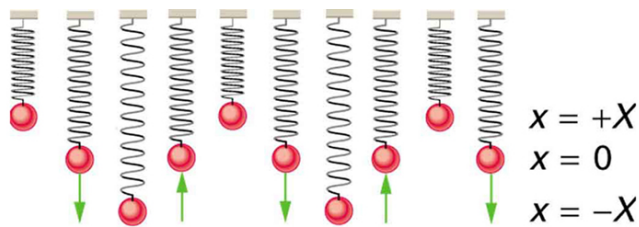
where $v_{\max} = 2\pi X/T = X\sqrt{k/m}$. The object has zero velocity at maximum displacement—for example, $v = 0$ when $t = 0$, and at that time $x = X$. The minus sign in the first equation for $v(t)$ gives the correct direction for the velocity. Just after the start of the motion, for instance, the velocity is negative because the system is moving back toward the equilibrium point. Finally, we can get an expression for acceleration using Newton's second law. [Then we have $x(t)$, $v(t)$, t , and $a(t)$, the quantities needed for kinematics and a description of simple harmonic motion.] According to Newton's second law, the acceleration is $a = F/m = kx/m$. So, $a(t)$ is also a cosine function:

Equation:

$$a(t) = -\frac{kX}{m} \cos \frac{2\pi t}{T}.$$

Hence, $a(t)$ is directly proportional to and in the opposite direction to $x(t)$.

[\[link\]](#) shows the simple harmonic motion of an object on a spring and presents graphs of $x(t)$, $v(t)$, and $a(t)$ versus time.



Graphs of $x(t)$, $v(t)$, and $a(t)$ versus t for the motion of an object on a spring. The net force on the object can be described by Hooke's law, and so the object undergoes simple harmonic motion. Note that the initial position has the vertical displacement at its maximum value X ; v is initially zero and then negative as the object moves down; and the initial acceleration

is negative, back toward the equilibrium position and becomes zero at that point.

The most important point here is that these equations are mathematically straightforward and are valid for all simple harmonic motion. They are very useful in visualizing waves associated with simple harmonic motion, including visualizing how waves add with one another.

Exercise:

Check Your Understanding

Problem:

Suppose you pluck a banjo string. You hear a single note that starts out loud and slowly quiets over time. Describe what happens to the sound waves in terms of period, frequency and amplitude as the sound decreases in volume.

Solution:

Frequency and period remain essentially unchanged. Only amplitude decreases as volume decreases.

Exercise:

Check Your Understanding

Problem:

A babysitter is pushing a child on a swing. At the point where the swing reaches x , where would the corresponding point on a wave of this motion be located?

Solution:

x is the maximum deformation, which corresponds to the amplitude of the wave. The point on the wave would either be at the very top or the very bottom of the curve.

Note:**PhET Explorations: Masses and Springs**

A realistic mass and spring laboratory. Hang masses from springs and adjust the spring stiffness and damping. You can even slow time. Transport the lab to different planets. A chart shows the kinetic, potential, and thermal energy for each spring.

https://phet.colorado.edu/sims/mass-spring-lab/mass-spring-lab_en.html

Section Summary

- Simple harmonic motion is oscillatory motion for a system that can be described only by Hooke's law. Such a system is also called a simple harmonic oscillator.
- Maximum displacement is the amplitude X . The period T and frequency f of a simple harmonic oscillator are given by

$T = 2\pi\sqrt{\frac{m}{k}}$ and $f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$, where m is the mass of the system.

- Displacement in simple harmonic motion as a function of time is given by $x(t) = X \cos \frac{2\pi t}{T}$.
- The velocity is given by $v(t) = -v_{\max} \sin \frac{2\pi t}{T}$, where $v_{\max} = \sqrt{k/m}X$.
- The acceleration is found to be $a(t) = -\frac{kX}{m} \cos \frac{2\pi t}{T}$.

Conceptual Questions

Exercise:**Problem:**

What conditions must be met to produce simple harmonic motion?

Exercise:

Problem:

- (a) If frequency is not constant for some oscillation, can the oscillation be simple harmonic motion?
- (b) Can you think of any examples of harmonic motion where the frequency may depend on the amplitude?

Exercise:**Problem:**

Give an example of a simple harmonic oscillator, specifically noting how its frequency is independent of amplitude.

Exercise:**Problem:**

Explain why you expect an object made of a stiff material to vibrate at a higher frequency than a similar object made of a spongy material.

Exercise:**Problem:**

As you pass a freight truck with a trailer on a highway, you notice that its trailer is bouncing up and down slowly. Is it more likely that the trailer is heavily loaded or nearly empty? Explain your answer.

Exercise:**Problem:**

Some people modify cars to be much closer to the ground than when manufactured. Should they install stiffer springs? Explain your answer.

Problems & Exercises**Exercise:**

Problem:

A type of cuckoo clock keeps time by having a mass bouncing on a spring, usually something cute like a cherub in a chair. What force constant is needed to produce a period of 0.500 s for a 0.0150-kg mass?

Solution:

2.37 N/m

Exercise:**Problem:**

If the spring constant of a simple harmonic oscillator is doubled, by what factor will the mass of the system need to change in order for the frequency of the motion to remain the same?

Exercise:**Problem:**

A 0.500-kg mass suspended from a spring oscillates with a period of 1.50 s. How much mass must be added to the object to change the period to 2.00 s?

Solution:

0.389 kg

Exercise:**Problem:**

By how much leeway (both percentage and mass) would you have in the selection of the mass of the object in the previous problem if you did not wish the new period to be greater than 2.01 s or less than 1.99 s?

Exercise:

Problem:

Suppose you attach the object with mass m to a vertical spring originally at rest, and let it bounce up and down. You release the object from rest at the spring's original rest length. (a) Show that the spring exerts an upward force of $2.00\ mg$ on the object at its lowest point. (b) If the spring has a force constant of $10.0\ \text{N/m}$ and a 0.25-kg -mass object is set in motion as described, find the amplitude of the oscillations. (c) Find the maximum velocity.

Exercise:**Problem:**

A diver on a diving board is undergoing simple harmonic motion. Her mass is $55.0\ \text{kg}$ and the period of her motion is $0.800\ \text{s}$. The next diver is a male whose period of simple harmonic oscillation is $1.05\ \text{s}$. What is his mass if the mass of the board is negligible?

Solution:

$94.7\ \text{kg}$

Exercise:**Problem:**

Suppose a diving board with no one on it bounces up and down in a simple harmonic motion with a frequency of $4.00\ \text{Hz}$. The board has an effective mass of $10.0\ \text{kg}$. What is the frequency of the simple harmonic motion of a 75.0-kg diver on the board?

Exercise:**Problem:**



This child's toy
relies on springs to
keep infants
entertained. (credit:
By Humboldtthead,
Flickr)

The device pictured in [\[link\]](#) entertains infants while keeping them from wandering. The child bounces in a harness suspended from a door frame by a spring constant.

(a) If the spring stretches 0.250 m while supporting an 8.0-kg child, what is its spring constant?

(b) What is the time for one complete bounce of this child? (c) What is the child's maximum velocity if the amplitude of her bounce is 0.200 m?

Exercise:

Problem:

A 90.0-kg skydiver hanging from a parachute bounces up and down with a period of 1.50 s. What is the new period of oscillation when a second skydiver, whose mass is 60.0 kg, hangs from the legs of the first, as seen in [\[link\]](#).



The oscillations of one skydiver are about to be affected by a second skydiver. (credit: U.S. Army, www.army.mil)

Solution:

1.94 s

Glossary

amplitude

the maximum displacement from the equilibrium position of an object oscillating around the equilibrium position

simple harmonic motion

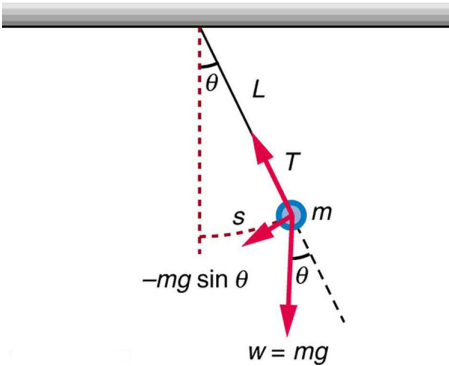
the oscillatory motion in a system where the net force can be described by Hooke's law

simple harmonic oscillator

a device that implements Hooke's law, such as a mass that is attached to a spring, with the other end of the spring being connected to a rigid support such as a wall

The Simple Pendulum

- Measure acceleration due to gravity.



A simple pendulum has a small-diameter bob and a string that has a very small mass but is strong enough not to stretch appreciably. The linear displacement from equilibrium is s , the length of the arc. Also shown are the forces on the bob, which result in a net force of $-mg \sin \theta$ toward the equilibrium position—that is, a restoring force.

Pendulums are in common usage. Some have crucial uses, such as in clocks; some are for fun, such as a child's swing; and some are just there, such as the sinker on a fishing line. For small displacements, a pendulum is a simple harmonic oscillator. A **simple pendulum** is defined to have an

object that has a small mass, also known as the pendulum bob, which is suspended from a light wire or string, such as shown in [\[link\]](#). Exploring the simple pendulum a bit further, we can discover the conditions under which it performs simple harmonic motion, and we can derive an interesting expression for its period.

We begin by defining the displacement to be the arc length s . We see from [\[link\]](#) that the net force on the bob is tangent to the arc and equals $-mg \sin \theta$. (The weight mg has components $mg \cos \theta$ along the string and $mg \sin \theta$ tangent to the arc.) Tension in the string exactly cancels the component $mg \cos \theta$ parallel to the string. This leaves a *net* restoring force back toward the equilibrium position at $\theta = 0$.

Now, if we can show that the restoring force is directly proportional to the displacement, then we have a simple harmonic oscillator. In trying to determine if we have a simple harmonic oscillator, we should note that for small angles (less than about 15°), $\sin \theta \approx \theta$ ($\sin \theta$ and θ differ by about 1% or less at smaller angles). Thus, for angles less than about 15° , the restoring force F is

Equation:

$$F \approx -mg\theta.$$

The displacement s is directly proportional to θ . When θ is expressed in radians, the arc length in a circle is related to its radius (L in this instance) by:

Equation:

$$s = L\theta,$$

so that

Equation:

$$\theta = \frac{s}{L}.$$

For small angles, then, the expression for the restoring force is:

Equation:

$$F \approx -\frac{mg}{L}s$$

This expression is of the form:

Equation:

$$F = -kx,$$

where the force constant is given by $k = mg/L$ and the displacement is given by $x = s$. For angles less than about 15° , the restoring force is directly proportional to the displacement, and the simple pendulum is a simple harmonic oscillator.

Using this equation, we can find the period of a pendulum for amplitudes less than about 15° . For the simple pendulum:

Equation:

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{m}{mg/L}}.$$

Thus,

Equation:

$$T = 2\pi\sqrt{\frac{L}{g}}$$

for the period of a simple pendulum. This result is interesting because of its simplicity. The only things that affect the period of a simple pendulum are its length and the acceleration due to gravity. The period is completely independent of other factors, such as mass. As with simple harmonic oscillators, the period T for a pendulum is nearly independent of amplitude,

especially if θ is less than about 15° . Even simple pendulum clocks can be finely adjusted and accurate.

Note the dependence of T on g . If the length of a pendulum is precisely known, it can actually be used to measure the acceleration due to gravity. Consider the following example.

Example:

Measuring Acceleration due to Gravity: The Period of a Pendulum

What is the acceleration due to gravity in a region where a simple pendulum having a length 75.000 cm has a period of 1.7357 s?

Strategy

We are asked to find g given the period T and the length L of a pendulum.

We can solve $T = 2\pi\sqrt{\frac{L}{g}}$ for g , assuming only that the angle of deflection is less than 15° .

Solution

1. Square $T = 2\pi\sqrt{\frac{L}{g}}$ and solve for g :

Equation:

$$g = 4\pi^2 \frac{L}{T^2}.$$

2. Substitute known values into the new equation:

Equation:

$$g = 4\pi^2 \frac{0.75000 \text{ m}}{(1.7357 \text{ s})^2}.$$

3. Calculate to find g :

Equation:

$$g = 9.8281 \text{ m/s}^2.$$

Discussion

This method for determining g can be very accurate. This is why length and period are given to five digits in this example. For the precision of the approximation $\sin \theta \approx \theta$ to be better than the precision of the pendulum length and period, the maximum displacement angle should be kept below about 0.5° .

Note:**Making Career Connections**

Knowing g can be important in geological exploration; for example, a map of g over large geographical regions aids the study of plate tectonics and helps in the search for oil fields and large mineral deposits.

Note:**Take Home Experiment: Determining g**

Use a simple pendulum to determine the acceleration due to gravity g in your own locale. Cut a piece of a string or dental floss so that it is about 1 m long. Attach a small object of high density to the end of the string (for example, a metal nut or a car key). Starting at an angle of less than 10° , allow the pendulum to swing and measure the pendulum's period for 10 oscillations using a stopwatch. Calculate g . How accurate is this measurement? How might it be improved?

Exercise:**Check Your Understanding**

Problem:

An engineer builds two simple pendula. Both are suspended from small wires secured to the ceiling of a room. Each pendulum hovers 2 cm above the floor. Pendulum 1 has a bob with a mass of 10 kg. Pendulum 2 has a bob with a mass of 100 kg. Describe how the motion of the pendula will differ if the bobs are both displaced by 12° .

Solution:

The movement of the pendula will not differ at all because the mass of the bob has no effect on the motion of a simple pendulum. The pendula are only affected by the period (which is related to the pendulum's length) and by the acceleration due to gravity.

Note:**PhET Explorations: Pendulum Lab**

Play with one or two pendulums and discover how the period of a simple pendulum depends on the length of the string, the mass of the pendulum bob, and the amplitude of the swing. It's easy to measure the period using the photogate timer. You can vary friction and the strength of gravity. Use the pendulum to find the value of g on planet X. Notice the anharmonic behavior at large amplitude.

https://phet.colorado.edu/sims/pendulum-lab/pendulum-lab_en.html

Section Summary

- A mass m suspended by a wire of length L is a simple pendulum and undergoes simple harmonic motion for amplitudes less than about 15° .

The period of a simple pendulum is

Equation:

$$T = 2\pi\sqrt{\frac{L}{g}},$$

where L is the length of the string and g is the acceleration due to gravity.

Conceptual Questions

Exercise:

Problem:

Pendulum clocks are made to run at the correct rate by adjusting the pendulum's length. Suppose you move from one city to another where the acceleration due to gravity is slightly greater, taking your pendulum clock with you, will you have to lengthen or shorten the pendulum to keep the correct time, other factors remaining constant? Explain your answer.

Problems & Exercises

As usual, the acceleration due to gravity in these problems is taken to be $g = 9.80 \text{ m/s}^2$, unless otherwise specified.

Exercise:

Problem:

What is the length of a pendulum that has a period of 0.500 s?

Solution:

6.21 cm

Exercise:

Problem:

Some people think a pendulum with a period of 1.00 s can be driven with “mental energy” or psycho kinetically, because its period is the same as an average heartbeat. True or not, what is the length of such a pendulum?

Exercise:

Problem: What is the period of a 1.00-m-long pendulum?

Solution:

2.01 s

Exercise:**Problem:**

How long does it take a child on a swing to complete one swing if her center of gravity is 4.00 m below the pivot?

Exercise:**Problem:**

The pendulum on a cuckoo clock is 5.00 cm long. What is its frequency?

Solution:

2.23 Hz

Exercise:**Problem:**

Two parakeets sit on a swing with their combined center of mass 10.0 cm below the pivot. At what frequency do they swing?

Exercise:

Problem:

(a) A pendulum that has a period of 3.00000 s and that is located where the acceleration due to gravity is 9.79 m/s^2 is moved to a location where the acceleration due to gravity is 9.82 m/s^2 . What is its new period? (b) Explain why so many digits are needed in the value for the period, based on the relation between the period and the acceleration due to gravity.

Solution:

(a) 2.99541 s

(b) Since the period is related to the square root of the acceleration of gravity, when the acceleration changes by 1% the period changes by $(0.01)^2 = 0.01\%$ so it is necessary to have at least 4 digits after the decimal to see the changes.

Exercise:**Problem:**

A pendulum with a period of 2.00000 s in one location ($g = 9.80 \text{ m/s}^2$) is moved to a new location where the period is now 1.99796 s. What is the acceleration due to gravity at its new location?

Exercise:**Problem:**

(a) What is the effect on the period of a pendulum if you double its length?

(b) What is the effect on the period of a pendulum if you decrease its length by 5.00%?

Solution:

(a) Period increases by a factor of 1.41 ($\sqrt{2}$)

(b) Period decreases to 97.5% of old period

Exercise:

Problem:

Find the ratio of the new/old periods of a pendulum if the pendulum were transported from Earth to the Moon, where the acceleration due to gravity is 1.63 m/s^2 .

Exercise:

Problem:

At what rate will a pendulum clock run on the Moon, where the acceleration due to gravity is 1.63 m/s^2 , if it keeps time accurately on Earth? That is, find the time (in hours) it takes the clock's hour hand to make one revolution on the Moon.

Solution:

Slow by a factor of 2.45

Exercise:

Problem:

Suppose the length of a clock's pendulum is changed by 1.000%, exactly at noon one day. What time will it read 24.00 hours later, assuming it the pendulum has kept perfect time before the change? Note that there are two answers, and perform the calculation to four-digit precision.

Exercise:

Problem:

If a pendulum-driven clock gains 5.00 s/day, what fractional change in pendulum length must be made for it to keep perfect time?

Solution:

length must increase by 0.0116%.

Glossary

simple pendulum

an object with a small mass suspended from a light wire or string

Energy and the Simple Harmonic Oscillator

- Determine the maximum speed of an oscillating system.

To study the energy of a simple harmonic oscillator, we first consider all the forms of energy it can have. We know from [Hooke's Law: Stress and Strain Revisited](#) that the energy stored in the deformation of a simple harmonic oscillator is a form of potential energy given by:

Equation:

$$PE_{\text{el}} = \frac{1}{2}kx^2.$$

Because a simple harmonic oscillator has no dissipative forces, the other important form of energy is kinetic energy KE. Conservation of energy for these two forms is:

Equation:

$$KE + PE_{\text{el}} = \text{constant}$$

or

Equation:

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \text{constant}.$$

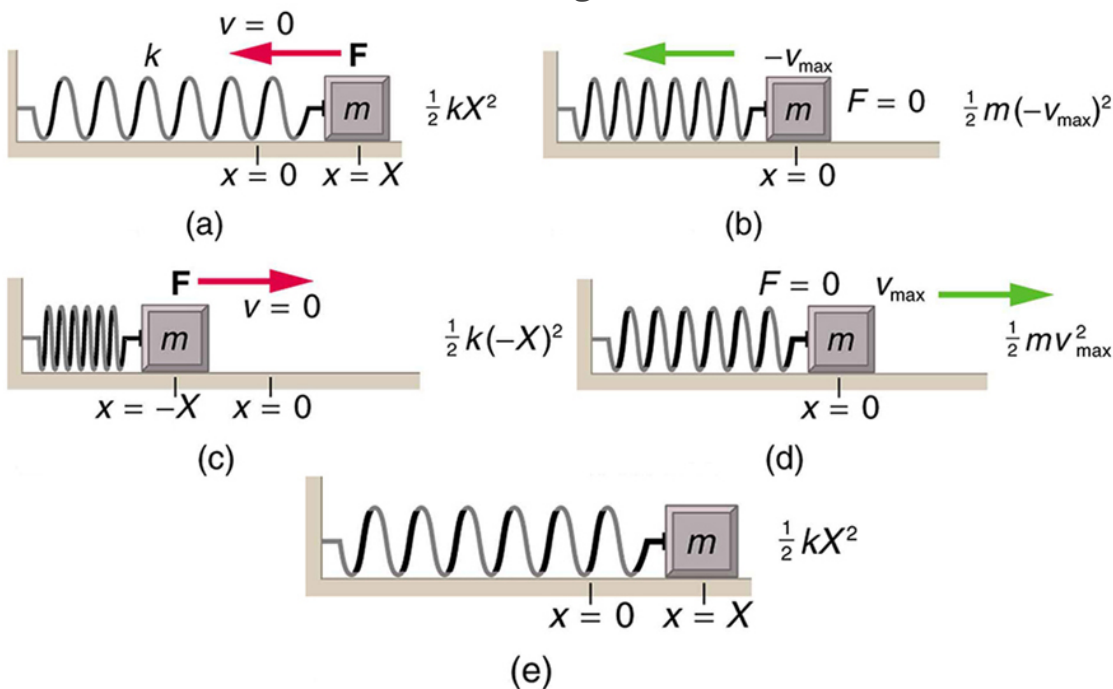
This statement of conservation of energy is valid for *all* simple harmonic oscillators, including ones where the gravitational force plays a role

Namely, for a simple pendulum we replace the velocity with $v = L\omega$, the spring constant with $k = mg/L$, and the displacement term with $x = L\theta$. Thus

Equation:

$$\frac{1}{2}mL^2\omega^2 + \frac{1}{2}mgL\theta^2 = \text{constant}.$$

In the case of undamped simple harmonic motion, the energy oscillates back and forth between kinetic and potential, going completely from one to the other as the system oscillates. So for the simple example of an object on a frictionless surface attached to a spring, as shown again in [\[link\]](#), the motion starts with all of the energy stored in the spring. As the object starts to move, the elastic potential energy is converted to kinetic energy, becoming entirely kinetic energy at the equilibrium position. It is then converted back into elastic potential energy by the spring, the velocity becomes zero when the kinetic energy is completely converted, and so on. This concept provides extra insight here and in later applications of simple harmonic motion, such as alternating current circuits.



The transformation of energy in simple harmonic motion is illustrated for an object attached to a spring on a frictionless surface.

The conservation of energy principle can be used to derive an expression for velocity v . If we start our simple harmonic motion with zero velocity and maximum displacement ($x = X$), then the total energy is

Equation:

$$\frac{1}{2}kX^2.$$

This total energy is constant and is shifted back and forth between kinetic energy and potential energy, at most times being shared by each. The conservation of energy for this system in equation form is thus:

Equation:

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{1}{2}kX^2.$$

Solving this equation for v yields:

Equation:

$$v = \pm \sqrt{\frac{k}{m}(X^2 - x^2)}.$$

Manipulating this expression algebraically gives:

Equation:

$$v = \pm \sqrt{\frac{k}{m}}X\sqrt{1 - \frac{x^2}{X^2}}$$

and so

Equation:

$$v = \pm v_{\max} \sqrt{1 - \frac{x^2}{X^2}},$$

where

Equation:

$$v_{\max} = \sqrt{\frac{k}{m}} X.$$

From this expression, we see that the velocity is a maximum (v_{\max}) at $x = 0$, as stated earlier in $v(t) = -v_{\max} \sin \frac{2\pi t}{T}$. Notice that the maximum velocity depends on three factors. Maximum velocity is directly proportional to amplitude. As you might guess, the greater the maximum displacement the greater the maximum velocity. Maximum velocity is also greater for stiffer systems, because they exert greater force for the same displacement. This observation is seen in the expression for v_{\max} ; it is proportional to the square root of the force constant k . Finally, the maximum velocity is smaller for objects that have larger masses, because the maximum velocity is inversely proportional to the square root of m . For a given force, objects that have large masses accelerate more slowly.

A similar calculation for the simple pendulum produces a similar result, namely:

Equation:

$$\omega_{\max} = \sqrt{\frac{g}{L}} \theta_{\max}.$$

Example:

Determine the Maximum Speed of an Oscillating System: A Bumpy Road

Suppose that a car is 900 kg and has a suspension system that has a force constant $k = 6.53 \times 10^4$ N/m. The car hits a bump and bounces with an amplitude of 0.100 m. What is its maximum vertical velocity if you assume no damping occurs?

Strategy

We can use the expression for v_{\max} given in $v_{\max} = \sqrt{\frac{k}{m}} X$ to determine the maximum vertical velocity. The variables m and k are given in the

problem statement, and the maximum displacement X is 0.100 m.

Solution

1. Identify known.

2. Substitute known values into $v_{\max} = \sqrt{\frac{k}{m}} X$:

Equation:

$$v_{\max} = \sqrt{\frac{6.53 \times 10^4 \text{ N/m}}{900 \text{ kg}}} (0.100 \text{ m}).$$

3. Calculate to find $v_{\max} = 0.852 \text{ m/s}$.

Discussion

This answer seems reasonable for a bouncing car. There are other ways to use conservation of energy to find v_{\max} . We could use it directly, as was done in the example featured in [Hooke's Law: Stress and Strain Revisited](#).

The small vertical displacement y of an oscillating simple pendulum, starting from its equilibrium position, is given as

Equation:

$$y(t) = a \sin \omega t,$$

where a is the amplitude, ω is the angular velocity and t is the time taken. Substituting $\omega = \frac{2\pi}{T}$, we have

Equation:

$$y(t) = a \sin \left(\frac{2\pi t}{T} \right).$$

Thus, the displacement of pendulum is a function of time as shown above. Also the velocity of the pendulum is given by

Equation:

$$v(t) = \frac{2a\pi}{T} \cos \left(\frac{2\pi t}{T} \right),$$

so the motion of the pendulum is a function of time.

Exercise:

Check Your Understanding

Problem:

Why does it hurt more if your hand is snapped with a ruler than with a loose spring, even if the displacement of each system is equal?

Solution:

The ruler is a stiffer system, which carries greater force for the same amount of displacement. The ruler snaps your hand with greater force, which hurts more.

Exercise:

Check Your Understanding

Problem:

You are observing a simple harmonic oscillator. Identify one way you could decrease the maximum velocity of the system.

Solution:

You could increase the mass of the object that is oscillating.

Section Summary

- Energy in the simple harmonic oscillator is shared between elastic potential energy and kinetic energy, with the total being constant:

Equation:

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \text{constant}.$$

- Maximum velocity depends on three factors: it is directly proportional to amplitude, it is greater for stiffer systems, and it is smaller for objects that have larger masses:

Equation:

$$v_{\max} = \sqrt{\frac{k}{m}} X.$$

Conceptual Questions

Exercise:

Problem:

Explain in terms of energy how dissipative forces such as friction reduce the amplitude of a harmonic oscillator. Also explain how a driving mechanism can compensate. (A pendulum clock is such a system.)

Problems & Exercises

Exercise:

Problem:

The length of nylon rope from which a mountain climber is suspended has a force constant of $1.40 \times 10^4 \text{ N/m}$.

- (a) What is the frequency at which he bounces, given his mass plus and the mass of his equipment are 90.0 kg?
- (b) How much would this rope stretch to break the climber's fall if he free-falls 2.00 m before the rope runs out of slack? Hint: Use conservation of energy.
- (c) Repeat both parts of this problem in the situation where twice this length of nylon rope is used.

Solution:

(a) 1.99 Hz

(b) 50.2 cm

(c) 1.41 Hz, 0.710 m

Exercise:**Problem: Engineering Application**

Near the top of the Citigroup Center building in New York City, there is an object with mass of 4.00×10^5 kg on springs that have adjustable force constants. Its function is to dampen wind-driven oscillations of the building by oscillating at the same frequency as the building is being driven—the driving force is transferred to the object, which oscillates instead of the entire building. (a) What effective force constant should the springs have to make the object oscillate with a period of 2.00 s? (b) What energy is stored in the springs for a 2.00-m displacement from equilibrium?

Solution:

(a) 3.95×10^6 N/m

(b) 7.90×10^6 J

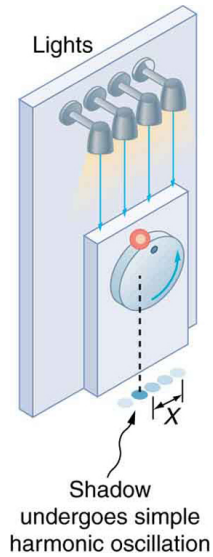
Uniform Circular Motion and Simple Harmonic Motion

- Compare simple harmonic motion with uniform circular motion.



The horses on this merry-go-round exhibit uniform circular motion. (credit: Wonderlane, Flickr)

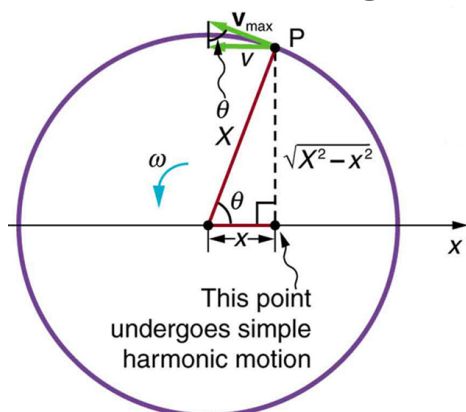
There is an easy way to produce simple harmonic motion by using uniform circular motion. [\[link\]](#) shows one way of using this method. A ball is attached to a uniformly rotating vertical turntable, and its shadow is projected on the floor as shown. The shadow undergoes simple harmonic motion. Hooke's law usually describes uniform circular motions (ω constant) rather than systems that have large visible displacements. So observing the projection of uniform circular motion, as in [\[link\]](#), is often easier than observing a precise large-scale simple harmonic oscillator. If studied in sufficient depth, simple harmonic motion produced in this manner can give considerable insight into many aspects of oscillations and waves and is very useful mathematically. In our brief treatment, we shall indicate some of the major features of this relationship and how they might be useful.



The shadow of a ball rotating at constant angular velocity ω on a turntable goes back and forth in precise simple harmonic motion.

[\[link\]](#) shows the basic relationship between uniform circular motion and simple harmonic motion. The point P travels around the circle at constant angular velocity ω . The point P is analogous to an object on the merry-go-

round. The projection of the position of P onto a fixed axis undergoes simple harmonic motion and is analogous to the shadow of the object. At the time shown in the figure, the projection has position x and moves to the left with velocity v . The velocity of the point P around the circle equals v_{\max} . The projection of v_{\max} on the x -axis is the velocity v of the simple harmonic motion along the x -axis.



A point P moving on a circular path with a constant angular velocity ω is undergoing uniform circular motion. Its projection on the x -axis undergoes simple harmonic motion. Also shown is the velocity of this point around the circle, v_{\max} , and its projection, which is v .

Note that these velocities form a similar triangle to the displacement triangle.

To see that the projection undergoes simple harmonic motion, note that its position x is given by

Equation:

$$x = X \cos \theta,$$

where $\theta = \omega t$, ω is the constant angular velocity, and X is the radius of the circular path. Thus,

Equation:

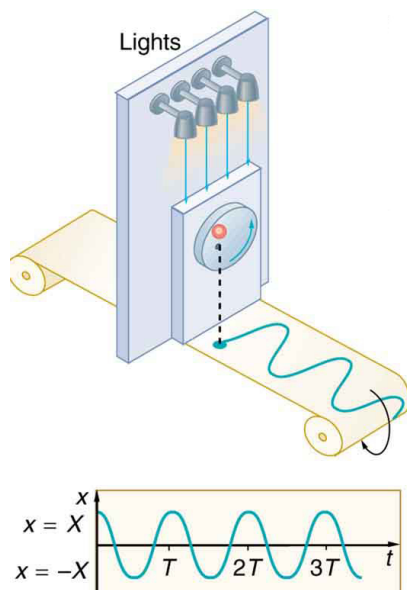
$$x = X \cos \omega t.$$

The angular velocity ω is in radians per unit time; in this case 2π radians is the time for one revolution T . That is, $\omega = 2\pi/T$. Substituting this expression for ω , we see that the position x is given by:

Equation:

$$x(t) = \cos\left(\frac{2\pi t}{T}\right).$$

This expression is the same one we had for the position of a simple harmonic oscillator in [Simple Harmonic Motion: A Special Periodic Motion](#). If we make a graph of position versus time as in [\[link\]](#), we see again the wavelike character (typical of simple harmonic motion) of the projection of uniform circular motion onto the x -axis.



The position of the projection of uniform circular motion performs simple harmonic motion, as this wavelike graph of x versus t indicates.

Now let us use [\[link\]](#) to do some further analysis of uniform circular motion as it relates to simple harmonic motion. The triangle formed by the velocities in the figure and the triangle formed by the displacements (X , x , and $\sqrt{X^2 - x^2}$) are similar right triangles. Taking ratios of similar sides, we see that

Equation:

$$\frac{v}{v_{\max}} = \frac{\sqrt{X^2 - x^2}}{X} = \sqrt{1 - \frac{x^2}{X^2}}.$$

We can solve this equation for the speed v or

Equation:

$$v = v_{\max} \sqrt{1 - \frac{x^2}{X^2}}.$$

This expression for the speed of a simple harmonic oscillator is exactly the same as the equation obtained from conservation of energy considerations in [Energy and the Simple Harmonic Oscillator](#). You can begin to see that it is possible to get all of the characteristics of simple harmonic motion from an analysis of the projection of uniform circular motion.

Finally, let us consider the period T of the motion of the projection. This period is the time it takes the point P to complete one revolution. That time is the circumference of the circle $2\pi X$ divided by the velocity around the circle, v_{\max} . Thus, the period T is

Equation:

$$T = \frac{2\pi X}{v_{\max}}.$$

We know from conservation of energy considerations that

Equation:

$$v_{\max} = \sqrt{\frac{k}{m}} X.$$

Solving this equation for X/v_{\max} gives

Equation:

$$\frac{X}{v_{\max}} = \sqrt{\frac{m}{k}}.$$

Substituting this expression into the equation for T yields

Equation:

$$T = 2\pi\sqrt{\frac{m}{k}}.$$

Thus, the period of the motion is the same as for a simple harmonic oscillator. We have determined the period for any simple harmonic oscillator using the relationship between uniform circular motion and simple harmonic motion.

Some modules occasionally refer to the connection between uniform circular motion and simple harmonic motion. Moreover, if you carry your study of physics and its applications to greater depths, you will find this relationship useful. It can, for example, help to analyze how waves add when they are superimposed.

Exercise:**Check Your Understanding****Problem:**

Identify an object that undergoes uniform circular motion. Describe how you could trace the simple harmonic motion of this object as a wave.

Solution:

A record player undergoes uniform circular motion. You could attach a dowel rod to one point on the outside edge of the turntable and attach a pen to the other end of the dowel. As the record player turns, the pen will move. You can drag a long piece of paper under the pen, capturing its motion as a wave.

Section Summary

A projection of uniform circular motion undergoes simple harmonic oscillation.

Problems & Exercises

Exercise:

Problem:

(a) What is the maximum velocity of an 85.0-kg person bouncing on a bathroom scale having a force constant of 1.50×10^6 N/m, if the amplitude of the bounce is 0.200 cm? (b) What is the maximum energy stored in the spring?

Solution:

a). 0.266 m/s

b). 3.00 J

Exercise:

Problem:

A novelty clock has a 0.0100-kg mass object bouncing on a spring that has a force constant of 1.25 N/m. What is the maximum velocity of the object if the object bounces 3.00 cm above and below its equilibrium position? (b) How many joules of kinetic energy does the object have at its maximum velocity?

Exercise:

Problem:

At what positions is the speed of a simple harmonic oscillator half its maximum? That is, what values of x/X give $v = \pm v_{\max}/2$, where X is the amplitude of the motion?

Solution:

$$\pm \frac{\sqrt{3}}{2}$$

Exercise:

Problem:

A ladybug sits 12.0 cm from the center of a Beatles music album spinning at 33.33 rpm. What is the maximum velocity of its shadow on the wall behind the turntable, if illuminated parallel to the record by the parallel rays of the setting Sun?

Damped Harmonic Motion

- Compare and discuss underdamped and overdamped oscillating systems.
- Explain critically damped system.



In order to counteract dampening forces, this mom needs to keep pushing the swing. (credit: Mohd Fazlin Mohd Effendy Ooi, Flickr)

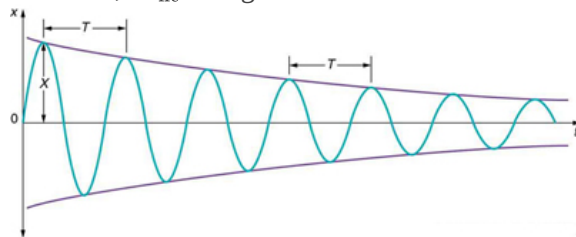
A guitar string stops oscillating a few seconds after being plucked. To keep a child happy on a swing, you must keep pushing. Although we can often make friction and other non-conservative forces negligibly small, completely undamped motion is rare. In fact, we may even want to damp oscillations, such as with car shock absorbers.

For a system that has a small amount of damping, the period and frequency are nearly the same as for simple harmonic motion, but the amplitude gradually decreases as shown in [\[link\]](#). This occurs because the non-conservative damping force removes energy from the system, usually in the form of thermal energy. In general, energy removal by non-conservative forces is described as

Equation:

$$W_{nc} = \Delta(KE + PE),$$

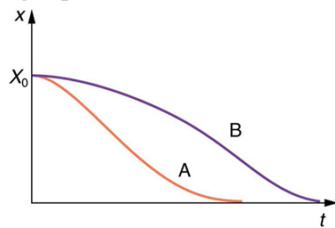
where W_{nc} is work done by a non-conservative force (here the damping force). For a damped harmonic oscillator, W_{nc} is negative because it removes mechanical energy ($KE + PE$) from the system.



In this graph of displacement versus time for a harmonic oscillator with a small amount of damping, the amplitude slowly decreases, but the period and frequency are nearly the

same as if the system were completely undamped.

If you gradually *increase* the amount of damping in a system, the period and frequency begin to be affected, because damping opposes and hence slows the back and forth motion. (The net force is smaller in both directions.) If there is very large damping, the system does not even oscillate—it slowly moves toward equilibrium. [\[link\]](#) shows the displacement of a harmonic oscillator for different amounts of damping. When we want to damp out oscillations, such as in the suspension of a car, we may want the system to return to equilibrium as quickly as possible. **Critical damping** is defined as the condition in which the damping of an oscillator results in it returning as quickly as possible to its equilibrium position. The critically damped system may overshoot the equilibrium position, but if it does, it will do so only once. Critical damping is represented by Curve A in [\[link\]](#). With less-than critical damping, the system will return to equilibrium faster but will overshoot and cross over one or more times. Such a system is **underdamped**; its displacement is represented by the curve in [\[link\]](#). Curve B in [\[link\]](#) represents an **overdamped** system. As with critical damping, it too may overshoot the equilibrium position, but will reach equilibrium over a longer period of time.



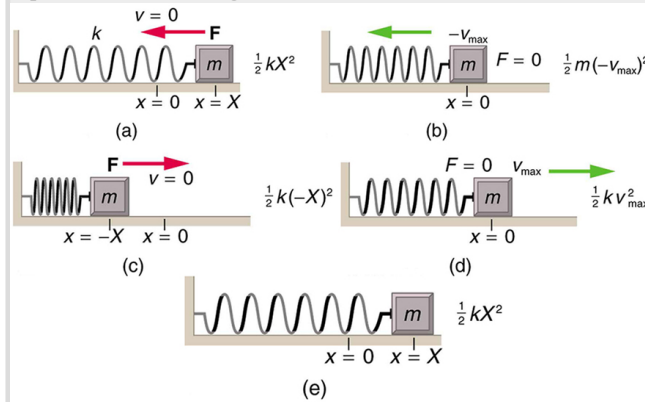
Displacement versus time
for a critically damped
harmonic oscillator (A)
and an overdamped
harmonic oscillator (B).
The critically damped
oscillator returns to
equilibrium at $X = 0$ in
the smallest time possible
without overshooting.

Critical damping is often desired, because such a system returns to equilibrium rapidly and remains at equilibrium as well. In addition, a constant force applied to a critically damped system moves the system to a new equilibrium position in the shortest time possible without overshooting or oscillating about the new position. For example, when you stand on bathroom scales that have a needle gauge, the needle moves to its equilibrium position without oscillating. It would be quite inconvenient if the needle oscillated about the new equilibrium position for a long time before settling. Damping forces can vary greatly in character. Friction, for example, is sometimes independent of velocity (as assumed in most places in this text). But many damping forces depend on velocity—sometimes in complex ways, sometimes simply being proportional to velocity.

Example:

Damping an Oscillatory Motion: Friction on an Object Connected to a Spring

Damping oscillatory motion is important in many systems, and the ability to control the damping is even more so. This is generally attained using non-conservative forces such as the friction between surfaces, and viscosity for objects moving through fluids. The following example considers friction. Suppose a 0.200-kg object is connected to a spring as shown in [\[link\]](#), but there is simple friction between the object and the surface, and the coefficient of friction μ_k is equal to 0.0800. (a) What is the frictional force between the surfaces? (b) What total distance does the object travel if it is released 0.100 m from equilibrium, starting at $v = 0$? The force constant of the spring is $k = 50.0 \text{ N/m}$.



The transformation of energy in simple harmonic motion is illustrated for an object attached to a spring on a frictionless surface.

Strategy

This problem requires you to integrate your knowledge of various concepts regarding waves, oscillations, and damping. To solve an integrated concept problem, you must first identify the physical principles involved. Part (a) is about the frictional force. This is a topic involving the application of Newton's Laws. Part (b) requires an understanding of work and conservation of energy, as well as some understanding of horizontal oscillatory systems.

Now that we have identified the principles we must apply in order to solve the problems, we need to identify the knowns and unknowns for each part of the question, as well as the quantity that is constant in Part (a) and Part (b) of the question.

Solution a

1. Choose the proper equation: Friction is $f = \mu_k mg$.
2. Identify the known values.
3. Enter the known values into the equation:

Equation:

$$f = (0.0800)(0.200 \text{ kg})(9.80 \text{ m/s}^2).$$

4. Calculate and convert units: $f = 0.157 \text{ N}$.

Discussion a

The force here is small because the system and the coefficients are small.

Solution b

Identify the known:

- The system involves elastic potential energy as the spring compresses and expands, friction that is related to the work done, and the kinetic energy as the body speeds up and slows down.
- Energy is not conserved as the mass oscillates because friction is a non-conservative force.
- The motion is horizontal, so gravitational potential energy does not need to be considered.
- Because the motion starts from rest, the energy in the system is initially $PE_{el,i} = (1/2)kX^2$. This energy is removed by work done by friction $W_{nc} = -fd$, where d is the total distance traveled and $f = \mu_k mg$ is the force of friction. When the system stops moving, the friction force will balance the force exerted by the spring, so $PE_{el,f} = (1/2)kx^2$ where x is the final position and is given by

Equation:

$$\begin{aligned} F_{el} &= f \\ kx &= \mu_k mg \\ x &= \frac{\mu_k mg}{k} \end{aligned}$$

1. By equating the work done to the energy removed, solve for the distance d .
2. The work done by the non-conservative forces equals the initial, stored elastic potential energy. Identify the correct equation to use:

Equation:

$$W_{nc} = \Delta(K.E + P.E) = PE_{el,f} - PE_{el,i} = \frac{1}{2}k\left(\left(\frac{\mu_k mg}{k}\right)^2 - X^2\right).$$

3. Recall that $W_{nc} = -fd$.
4. Enter the friction as $f = \mu_k mg$ into $W_{nc} = -fd$, thus

Equation:

$$W_{nc} = -\mu_k mgd.$$

5. Combine these two equations to find

Equation:

$$\frac{1}{2}k\left(\left(\frac{\mu_k mg}{k}\right)^2 - X^2\right) = -\mu_k mgd.$$

6. Solve the equation for d :

Equation:

$$d = \frac{k}{2\mu_k mg} \left(X^2 - \left(\frac{\mu_k mg}{k} \right)^2 \right).$$

7. Enter the known values into the resulting equation:

Equation:

$$d = \frac{50.0 \text{ N/m}}{2(0.0800)(0.200 \text{ kg})(9.80 \text{ m/s}^2)} \left((0.100 \text{ m})^2 - \frac{(0.0800)(0.200 \text{ kg})(9.80 \text{ m/s}^2)^2}{50.0 \text{ N/m}} \right).$$

8. Calculate d and convert units:

Equation:

$$d = 1.59 \text{ m}.$$

Discussion b

This is the total distance traveled back and forth across $x = 0$, which is the undamped equilibrium position. The number of oscillations about the equilibrium position will be more than $d/X = (1.59 \text{ m})/(0.100 \text{ m}) = 15.9$ because the amplitude of the oscillations is decreasing with time. At the end of the motion, this system will not return to $x = 0$ for this type of damping force, because static friction will exceed the restoring force. This system is underdamped. In contrast, an overdamped system with a simple constant damping force would not cross the equilibrium position $x = 0$ a single time. For example, if this system had a damping force 20 times greater, it would only move 0.0484 m toward the equilibrium position from its original 0.100-m position.

This worked example illustrates how to apply problem-solving strategies to situations that integrate the different concepts you have learned. The first step is to identify the physical principles involved in the problem. The second step is to solve for the unknowns using familiar problem-solving strategies. These are found throughout the text, and many worked examples show how to use them for single topics. In this integrated concepts example, you can see how to apply them across several topics. You will find these techniques useful in applications of physics outside a physics course, such as in your profession, in other science disciplines, and in everyday life.

Exercise:**Check Your Understanding**

Problem: Why are completely undamped harmonic oscillators so rare?

Solution:

Friction often comes into play whenever an object is moving. Friction causes damping in a harmonic oscillator.

Exercise:**Check Your Understanding**

Problem: Describe the difference between overdamping, underdamping, and critical damping.

Solution:

An overdamped system moves slowly toward equilibrium. An underdamped system moves quickly to equilibrium, but will oscillate about the equilibrium point as it does so. A critically damped system moves as quickly as possible toward equilibrium without oscillating about the equilibrium.

Section Summary

- Damped harmonic oscillators have non-conservative forces that dissipate their energy.
- Critical damping returns the system to equilibrium as fast as possible without overshooting.
- An underdamped system will oscillate through the equilibrium position.
- An overdamped system moves more slowly toward equilibrium than one that is critically damped.

Conceptual Questions**Exercise:**

Problem:

Give an example of a damped harmonic oscillator. (They are more common than undamped or simple harmonic oscillators.)

Exercise:

Problem: How would a car bounce after a bump under each of these conditions?

- overdamping
- underdamping
- critical damping

Exercise:**Problem:**

Most harmonic oscillators are damped and, if undriven, eventually come to a stop. How is this observation related to the second law of thermodynamics?

Problems & Exercises**Exercise:****Problem:**

The amplitude of a lightly damped oscillator decreases by 3.0% during each cycle. What percentage of the mechanical energy of the oscillator is lost in each cycle?

Glossary**critical damping**

the condition in which the damping of an oscillator causes it to return as quickly as possible to its equilibrium position without oscillating back and forth about this position

over damping

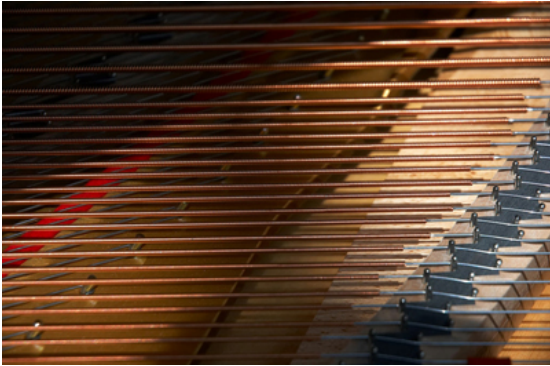
the condition in which damping of an oscillator causes it to return to equilibrium without oscillating; oscillator moves more slowly toward equilibrium than in the critically damped system

under damping

the condition in which damping of an oscillator causes it to return to equilibrium with the amplitude gradually decreasing to zero; system returns to equilibrium faster but overshoots and crosses the equilibrium position one or more times

Forced Oscillations and Resonance

- Observe resonance of a paddle ball on a string.
- Observe amplitude of a damped harmonic oscillator.

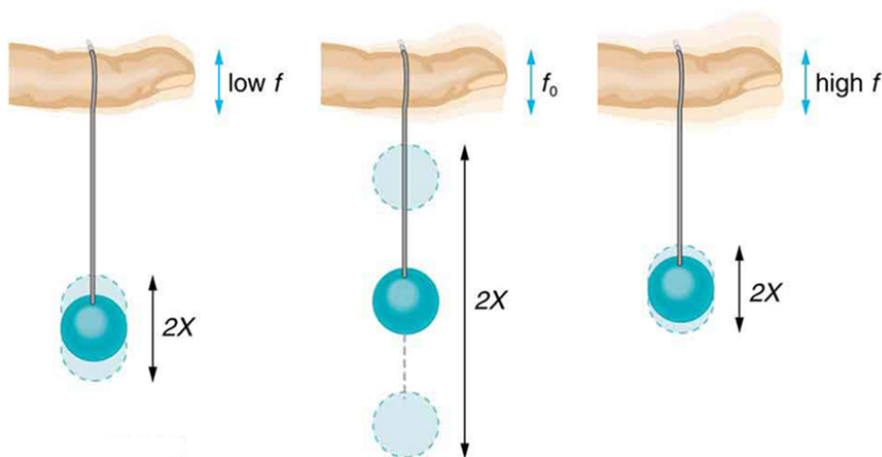


You can cause the strings in a piano to vibrate simply by producing sound waves from your voice. (credit: Matt Billings, Flickr)

Sit in front of a piano sometime and sing a loud brief note at it with the dampers off its strings. It will sing the same note back at you—the strings, having the same frequencies as your voice, are resonating in response to the forces from the sound waves that you sent to them. Your voice and a piano's strings is a good example of the fact that objects—in this case, piano strings—can be forced to oscillate but oscillate best at their natural frequency. In this section, we shall briefly explore applying a *periodic driving force* acting on a simple harmonic oscillator. The driving force puts energy into the system at a certain frequency, not necessarily the same as the natural frequency of the system. The **natural frequency** is the frequency at which a system would oscillate if there were no driving and no damping force.

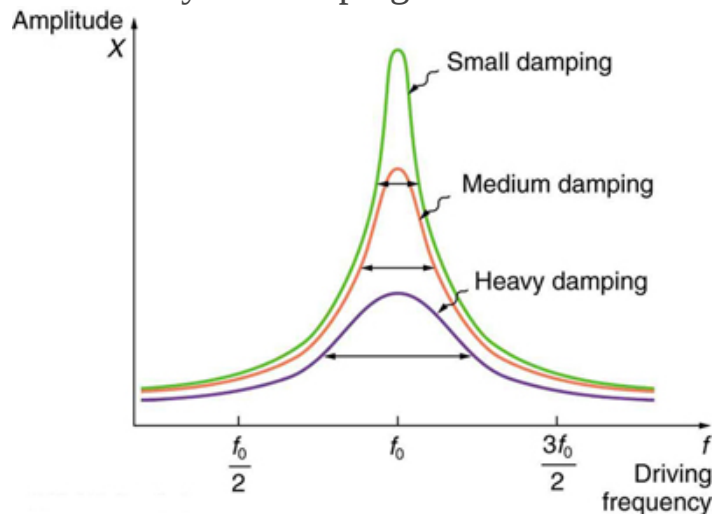
Most of us have played with toys involving an object supported on an elastic band, something like the paddle ball suspended from a finger in [\[link\]](#). Imagine the finger in the figure is your finger. At first you hold your

finger steady, and the ball bounces up and down with a small amount of damping. If you move your finger up and down slowly, the ball will follow along without bouncing much on its own. As you increase the frequency at which you move your finger up and down, the ball will respond by oscillating with increasing amplitude. When you drive the ball at its natural frequency, the ball's oscillations increase in amplitude with each oscillation for as long as you drive it. The phenomenon of driving a system with a frequency equal to its natural frequency is called **resonance**. A system being driven at its natural frequency is said to **resonate**. As the driving frequency gets progressively higher than the resonant or natural frequency, the amplitude of the oscillations becomes smaller, until the oscillations nearly disappear and your finger simply moves up and down with little effect on the ball.



The paddle ball on its rubber band moves in response to the finger supporting it. If the finger moves with the natural frequency f_0 of the ball on the rubber band, then a resonance is achieved, and the amplitude of the ball's oscillations increases dramatically. At higher and lower driving frequencies, energy is transferred to the ball less efficiently, and it responds with lower-amplitude oscillations.

[\[link\]](#) shows a graph of the amplitude of a damped harmonic oscillator as a function of the frequency of the periodic force driving it. There are three curves on the graph, each representing a different amount of damping. All three curves peak at the point where the frequency of the driving force equals the natural frequency of the harmonic oscillator. The highest peak, or greatest response, is for the least amount of damping, because less energy is removed by the damping force.



Amplitude of a harmonic oscillator as a function of the frequency of the driving force. The curves represent the same oscillator with the same natural frequency but with different amounts of damping. Resonance occurs when the driving frequency equals the natural frequency, and the greatest response is for the least amount of damping. The narrowest response is also for the least damping.

It is interesting that the widths of the resonance curves shown in [\[link\]](#) depend on damping: the less the damping, the narrower the resonance. The message is that if you want a driven oscillator to resonate at a very specific frequency, you need as little damping as possible. Little damping is the case

for piano strings and many other musical instruments. Conversely, if you want small-amplitude oscillations, such as in a car's suspension system, then you want heavy damping. Heavy damping reduces the amplitude, but the tradeoff is that the system responds at more frequencies.

These features of driven harmonic oscillators apply to a huge variety of systems. When you tune a radio, for example, you are adjusting its resonant frequency so that it only oscillates to the desired station's broadcast (driving) frequency. The more selective the radio is in discriminating between stations, the smaller its damping. Magnetic resonance imaging (MRI) is a widely used medical diagnostic tool in which atomic nuclei (mostly hydrogen nuclei) are made to resonate by incoming radio waves (on the order of 100 MHz). A child on a swing is driven by a parent at the swing's natural frequency to achieve maximum amplitude. In all of these cases, the efficiency of energy transfer from the driving force into the oscillator is best at resonance. Speed bumps and gravel roads prove that even a car's suspension system is not immune to resonance. In spite of finely engineered shock absorbers, which ordinarily convert mechanical energy to thermal energy almost as fast as it comes in, speed bumps still cause a large-amplitude oscillation. On gravel roads that are corrugated, you may have noticed that if you travel at the "wrong" speed, the bumps are very noticeable whereas at other speeds you may hardly feel the bumps at all. [\[link\]](#) shows a photograph of a famous example (the Tacoma Narrows Bridge) of the destructive effects of a driven harmonic oscillation. The Millennium Bridge in London was closed for a short period of time for the same reason while inspections were carried out.

In our bodies, the chest cavity is a clear example of a system at resonance. The diaphragm and chest wall drive the oscillations of the chest cavity which result in the lungs inflating and deflating. The system is critically damped and the muscular diaphragm oscillates at the resonant value for the system, making it highly efficient.



In 1940, the Tacoma Narrows Bridge in Washington state collapsed. Heavy cross winds drove the bridge into oscillations at its resonant frequency. Damping decreased when support cables broke loose and started to slip over the towers, allowing increasingly greater amplitudes until the structure failed (credit: PRI's *Studio 360*, via Flickr)

Exercise:
Check Your Understanding

Problem:

A famous magic trick involves a performer singing a note toward a crystal glass until the glass shatters. Explain why the trick works in terms of resonance and natural frequency.

Solution:

The performer must be singing a note that corresponds to the natural frequency of the glass. As the sound wave is directed at the glass, the glass responds by resonating at the same frequency as the sound wave.

With enough energy introduced into the system, the glass begins to vibrate and eventually shatters.

Section Summary

- A system's natural frequency is the frequency at which the system will oscillate if not affected by driving or damping forces.
- A periodic force driving a harmonic oscillator at its natural frequency produces resonance. The system is said to resonate.
- The less damping a system has, the higher the amplitude of the forced oscillations near resonance. The more damping a system has, the broader response it has to varying driving frequencies.

Conceptual Questions

Exercise:

Problem:

Why are soldiers in general ordered to “route step” (walk out of step) across a bridge?

Problems & Exercises

Exercise:

Problem:

How much energy must the shock absorbers of a 1200-kg car dissipate in order to damp a bounce that initially has a velocity of 0.800 m/s at the equilibrium position? Assume the car returns to its original vertical position.

Solution:

384 J

Exercise:**Problem:**

If a car has a suspension system with a force constant of $5.00 \times 10^4 \text{ N/m}$, how much energy must the car's shocks remove to dampen an oscillation starting with a maximum displacement of 0.0750 m?

Exercise:**Problem:**

(a) How much will a spring that has a force constant of 40.0 N/m be stretched by an object with a mass of 0.500 kg when hung motionless from the spring? (b) Calculate the decrease in gravitational potential energy of the 0.500-kg object when it descends this distance. (c) Part of this gravitational energy goes into the spring. Calculate the energy stored in the spring by this stretch, and compare it with the gravitational potential energy. Explain where the rest of the energy might go.

Solution:

(a). 0.123 m

(b). -0.600 J

(c). 0.300 J. The rest of the energy may go into heat caused by friction and other damping forces.

Exercise:

Problem:

Suppose you have a 0.750-kg object on a horizontal surface connected to a spring that has a force constant of 150 N/m. There is simple friction between the object and surface with a static coefficient of friction $\mu_s = 0.100$. (a) How far can the spring be stretched without moving the mass? (b) If the object is set into oscillation with an amplitude twice the distance found in part (a), and the kinetic coefficient of friction is $\mu_k = 0.0850$, what total distance does it travel before stopping? Assume it starts at the maximum amplitude.

Exercise:**Problem:**

Engineering Application: A suspension bridge oscillates with an effective force constant of 1.00×10^8 N/m. (a) How much energy is needed to make it oscillate with an amplitude of 0.100 m? (b) If soldiers march across the bridge with a cadence equal to the bridge's natural frequency and impart 1.00×10^4 J of energy each second, how long does it take for the bridge's oscillations to go from 0.100 m to 0.500 m amplitude?

Solution:

(a) 5.00×10^5 J

(b) 1.20×10^3 s

Glossary

natural frequency

the frequency at which a system would oscillate if there were no driving and no damping forces

resonance

the phenomenon of driving a system with a frequency equal to the system's natural frequency

resonate

a system being driven at its natural frequency

Waves

- State the characteristics of a wave.
- Calculate the velocity of wave propagation.



Waves in the ocean behave similarly to all other types of waves. (credit: Steve Jurveston, Flickr)

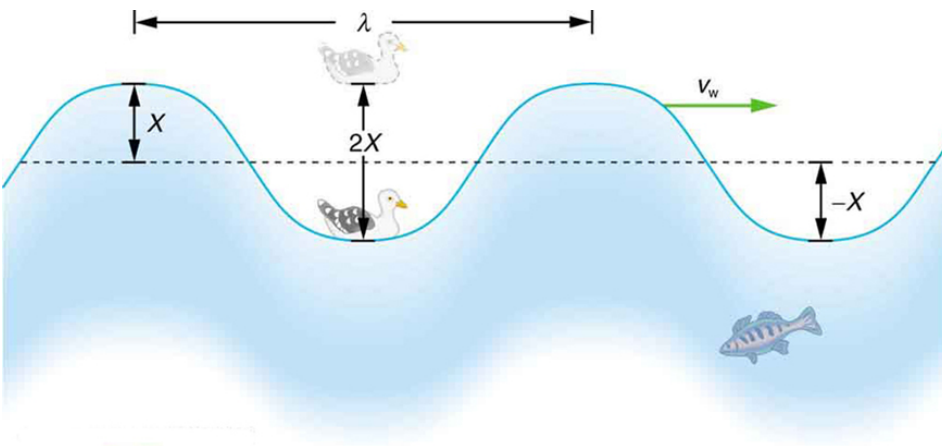
What do we mean when we say something is a wave? The most intuitive and easiest wave to imagine is the familiar water wave. More precisely, a **wave** is a disturbance that propagates, or moves from the place it was created. For water waves, the disturbance is in the surface of the water, perhaps created by a rock thrown into a pond or by a swimmer splashing the surface repeatedly. For sound waves, the disturbance is a change in air pressure, perhaps created by the oscillating cone inside a speaker. For earthquakes, there are several types of disturbances, including disturbance of Earth's surface and pressure disturbances under the surface. Even radio waves are most easily understood using an analogy with water waves. Visualizing water waves is useful because there is more to it than just a mental image. Water waves exhibit characteristics common to all waves, such as amplitude, period, frequency and energy. All wave characteristics can be described by a small set of underlying principles.

A wave is a disturbance that propagates, or moves from the place it was created. The simplest waves repeat themselves for several cycles and are associated with simple harmonic motion. Let us start by considering the simplified water wave in [\[link\]](#). The wave is an up and down disturbance of the water surface. It causes a sea gull to move up and down in simple harmonic motion as the wave crests and troughs (peaks and valleys) pass under the bird. The time for one complete up and down motion is the wave's period T . The wave's frequency is $f = 1/T$, as usual. The wave itself moves to the right in the figure. This movement of the wave is actually the disturbance moving to the right, not the water itself (or the bird would move to the right). We define **wave velocity** v_w to be the speed at which the disturbance moves. Wave velocity is sometimes also called the *propagation velocity* or *propagation speed*, because the disturbance propagates from one location to another.

Note:

Misconception Alert

Many people think that water waves push water from one direction to another. In fact, the particles of water tend to stay in one location, save for moving up and down due to the energy in the wave. The energy moves forward through the water, but the water stays in one place. If you feel yourself pushed in an ocean, what you feel is the energy of the wave, not a rush of water.



An idealized ocean wave passes under a sea gull that bobs up and down in simple harmonic motion. The wave has a wavelength λ , which is the distance between adjacent identical parts of the wave. The up and down disturbance of the surface propagates parallel to the surface at a speed v_w .

The water wave in the figure also has a length associated with it, called its **wavelength** λ , the distance between adjacent identical parts of a wave. (λ is the distance parallel to the direction of propagation.) The speed of propagation v_w is the distance the wave travels in a given time, which is one wavelength in the time of one period. In equation form, that is

Equation:

$$v_w = \frac{\lambda}{T}$$

or

Equation:

$$v_w = f\lambda.$$

This fundamental relationship holds for all types of waves. For water waves, v_w is the speed of a surface wave; for sound, v_w is the speed of sound; and for visible light, v_w is the speed of light, for example.

Note:

Take-Home Experiment: Waves in a Bowl

Fill a large bowl or basin with water and wait for the water to settle so there are no ripples. Gently drop a cork into the middle of the bowl. Estimate the wavelength and period of oscillation of the water wave that propagates away from the cork. Remove the cork from the bowl and wait

for the water to settle again. Gently drop the cork at a height that is different from the first drop. Does the wavelength depend upon how high above the water the cork is dropped?

Example:

Calculate the Velocity of Wave Propagation: Gull in the Ocean

Calculate the wave velocity of the ocean wave in [\[link\]](#) if the distance between wave crests is 10.0 m and the time for a sea gull to bob up and down is 5.00 s.

Strategy

We are asked to find v_w . The given information tells us that $\lambda = 10.0$ m and $T = 5.00$ s. Therefore, we can use $v_w = \frac{\lambda}{T}$ to find the wave velocity.

Solution

1. Enter the known values into $v_w = \frac{\lambda}{T}$:

Equation:

$$v_w = \frac{10.0 \text{ m}}{5.00 \text{ s}}.$$

2. Solve for v_w to find $v_w = 2.00$ m/s.

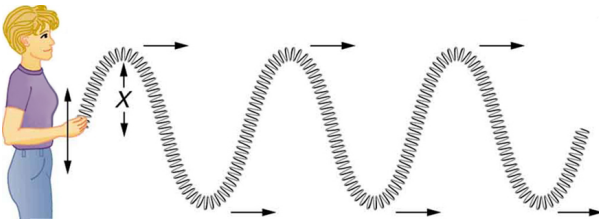
Discussion

This slow speed seems reasonable for an ocean wave. Note that the wave moves to the right in the figure at this speed, not the varying speed at which the sea gull moves up and down.

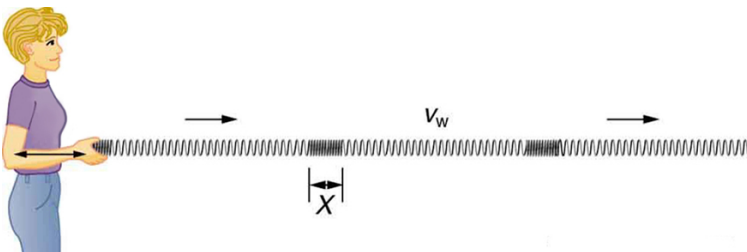
Transverse and Longitudinal Waves

A simple wave consists of a periodic disturbance that propagates from one place to another. The wave in [\[link\]](#) propagates in the horizontal direction while the surface is disturbed in the vertical direction. Such a wave is called a **transverse wave** or shear wave; in such a wave, the disturbance is perpendicular to the direction of propagation. In contrast, in a **longitudinal**

wave or compressional wave, the disturbance is parallel to the direction of propagation. [\[link\]](#) shows an example of a longitudinal wave. The size of the disturbance is its amplitude X and is completely independent of the speed of propagation v_w .



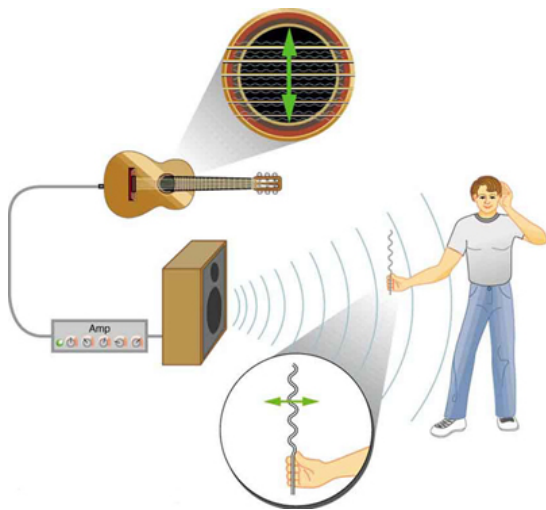
In this example of a transverse wave, the wave propagates horizontally, and the disturbance in the cord is in the vertical direction.



In this example of a longitudinal wave, the wave propagates horizontally, and the disturbance in the cord is also in the horizontal direction.

Waves may be transverse, longitudinal, or *a combination of the two*. (Water waves are actually a combination of transverse and longitudinal. The simplified water wave illustrated in [\[link\]](#) shows no longitudinal motion of the bird.) The waves on the strings of musical instruments are transverse—so are electromagnetic waves, such as visible light.

Sound waves in air and water are longitudinal. Their disturbances are periodic variations in pressure that are transmitted in fluids. Fluids do not have appreciable shear strength, and thus the sound waves in them must be longitudinal or compressional. Sound in solids can be both longitudinal and transverse.



The wave on a guitar string is transverse. The sound wave rattles a sheet of paper in a direction that shows the sound wave is longitudinal.

Earthquake waves under Earth's surface also have both longitudinal and transverse components (called compressional or P-waves and shear or S-waves, respectively). These components have important individual characteristics—they propagate at different speeds, for example.

Earthquakes also have surface waves that are similar to surface waves on water.

Exercise:

Check Your Understanding

Problem:

Why is it important to differentiate between longitudinal and transverse waves?

Solution:

In the different types of waves, energy can propagate in a different direction relative to the motion of the wave. This is important to understand how different types of waves affect the materials around them.

Note:

PhET Explorations: Wave on a String

Watch a string vibrate in slow motion. Wiggle the end of the string and make waves, or adjust the frequency and amplitude of an oscillator. Adjust the damping and tension. The end can be fixed, loose, or open.

https://phet.colorado.edu/sims/html/wave-on-a-string/latest/wave-on-a-string_en.html

Section Summary

- A wave is a disturbance that moves from the point of creation with a wave velocity v_w .
- A wave has a wavelength λ , which is the distance between adjacent identical parts of the wave.
- Wave velocity and wavelength are related to the wave's frequency and period by $v_w = \frac{\lambda}{T}$ or $v_w = f\lambda$.

- A transverse wave has a disturbance perpendicular to its direction of propagation, whereas a longitudinal wave has a disturbance parallel to its direction of propagation.

Conceptual Questions

Exercise:

Problem:

Give one example of a transverse wave and another of a longitudinal wave, being careful to note the relative directions of the disturbance and wave propagation in each.

Exercise:

Problem:

What is the difference between propagation speed and the frequency of a wave? Does one or both affect wavelength? If so, how?

Problems & Exercises

Exercise:

Problem:

Storms in the South Pacific can create waves that travel all the way to the California coast, which are 12,000 km away. How long does it take them if they travel at 15.0 m/s?

Solution:

Equation:

$$t = 9.26 \text{ d}$$

Exercise:

Problem:

Waves on a swimming pool propagate at 0.750 m/s. You splash the water at one end of the pool and observe the wave go to the opposite end, reflect, and return in 30.0 s. How far away is the other end of the pool?

Exercise:**Problem:**

Wind gusts create ripples on the ocean that have a wavelength of 5.00 cm and propagate at 2.00 m/s. What is their frequency?

Solution:**Equation:**

$$f = 40.0 \text{ Hz}$$

Exercise:**Problem:**

How many times a minute does a boat bob up and down on ocean waves that have a wavelength of 40.0 m and a propagation speed of 5.00 m/s?

Exercise:**Problem:**

Scouts at a camp shake the rope bridge they have just crossed and observe the wave crests to be 8.00 m apart. If they shake it the bridge twice per second, what is the propagation speed of the waves?

Solution:**Equation:**

$$v_w = 16.0 \text{ m/s}$$

Exercise:**Problem:**

What is the wavelength of the waves you create in a swimming pool if you splash your hand at a rate of 2.00 Hz and the waves propagate at 0.800 m/s?

Exercise:**Problem:**

What is the wavelength of an earthquake that shakes you with a frequency of 10.0 Hz and gets to another city 84.0 km away in 12.0 s?

Solution:**Equation:**

$$\lambda = 700 \text{ m}$$

Exercise:**Problem:**

Radio waves transmitted through space at $3.00 \times 10^8 \text{ m/s}$ by the Voyager spacecraft have a wavelength of 0.120 m. What is their frequency?

Exercise:**Problem:**

Your ear is capable of differentiating sounds that arrive at the ear just 1.00 ms apart. What is the minimum distance between two speakers that produce sounds that arrive at noticeably different times on a day when the speed of sound is 340 m/s?

Solution:**Equation:**

$$d = 34.0 \text{ cm}$$

Exercise:

Problem:

(a) Seismographs measure the arrival times of earthquakes with a precision of 0.100 s. To get the distance to the epicenter of the quake, they compare the arrival times of S- and P-waves, which travel at different speeds. ([link](#)) If S- and P-waves travel at 4.00 and 7.20 km/s, respectively, in the region considered, how precisely can the distance to the source of the earthquake be determined? (b) Seismic waves from underground detonations of nuclear bombs can be used to locate the test site and detect violations of test bans. Discuss whether your answer to (a) implies a serious limit to such detection. (Note also that the uncertainty is greater if there is an uncertainty in the propagation speeds of the S- and P-waves.)



A seismograph as described in above problem.(credit: Oleg Alexandrov)

Glossary

longitudinal wave

a wave in which the disturbance is parallel to the direction of propagation

transverse wave

a wave in which the disturbance is perpendicular to the direction of propagation

wave velocity

the speed at which the disturbance moves. Also called the propagation velocity or propagation speed

wavelength

the distance between adjacent identical parts of a wave

Superposition and Interference

- Explain standing waves.
- Describe the mathematical representation of overtones and beat frequency.



These waves result from the superposition of several waves from different sources, producing a complex pattern.
(credit: waterborough, Wikimedia Commons)

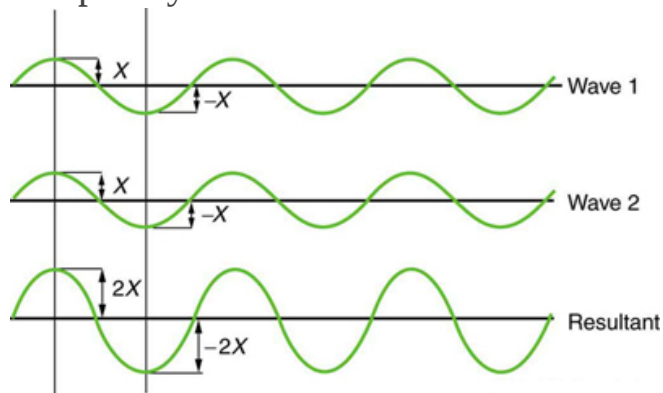
Most waves do not look very simple. They look more like the waves in [\[link\]](#) than like the simple water wave considered in [Waves](#). (Simple waves may be created by a simple harmonic oscillation, and thus have a sinusoidal shape). Complex waves are more interesting, even beautiful, but they look formidable. Most waves appear complex because they result from several simple waves adding together. Luckily, the rules for adding waves are quite simple.

When two or more waves arrive at the same point, they superimpose themselves on one another. More specifically, the disturbances of waves are superimposed when they come together—a phenomenon called **superposition**. Each disturbance corresponds to a force, and forces add. If the disturbances are along the same line, then the resulting wave is a simple

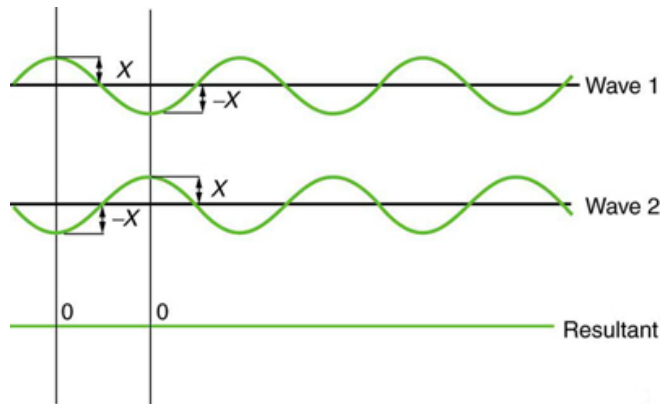
addition of the disturbances of the individual waves—that is, their amplitudes add. [\[link\]](#) and [\[link\]](#) illustrate superposition in two special cases, both of which produce simple results.

[\[link\]](#) shows two identical waves that arrive at the same point exactly in phase. The crests of the two waves are precisely aligned, as are the troughs. This superposition produces pure **constructive interference**. Because the disturbances add, pure constructive interference produces a wave that has twice the amplitude of the individual waves, but has the same wavelength.

[\[link\]](#) shows two identical waves that arrive exactly out of phase—that is, precisely aligned crest to trough—producing pure **destructive interference**. Because the disturbances are in the opposite direction for this superposition, the resulting amplitude is zero for pure destructive interference—the waves completely cancel.



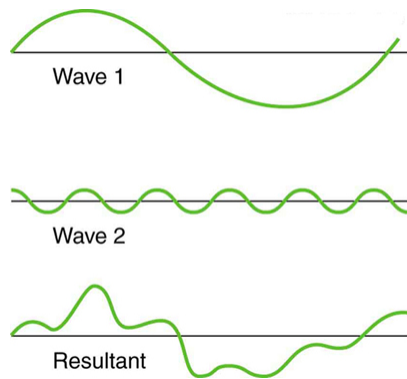
Pure constructive interference of two identical waves produces one with twice the amplitude, but the same wavelength.



Pure destructive interference of two identical waves produces zero amplitude, or complete cancellation.

While pure constructive and pure destructive interference do occur, they require precisely aligned identical waves. The superposition of most waves produces a combination of constructive and destructive interference and can vary from place to place and time to time. Sound from a stereo, for example, can be loud in one spot and quiet in another. Varying loudness means the sound waves add partially constructively and partially destructively at different locations. A stereo has at least two speakers creating sound waves, and waves can reflect from walls. All these waves superimpose. An example of sounds that vary over time from constructive to destructive is found in the combined whine of airplane jets heard by a stationary passenger. The combined sound can fluctuate up and down in volume as the sound from the two engines varies in time from constructive to destructive. These examples are of waves that are similar.

An example of the superposition of two dissimilar waves is shown in [\[link\]](#). Here again, the disturbances add and subtract, producing a more complicated looking wave.

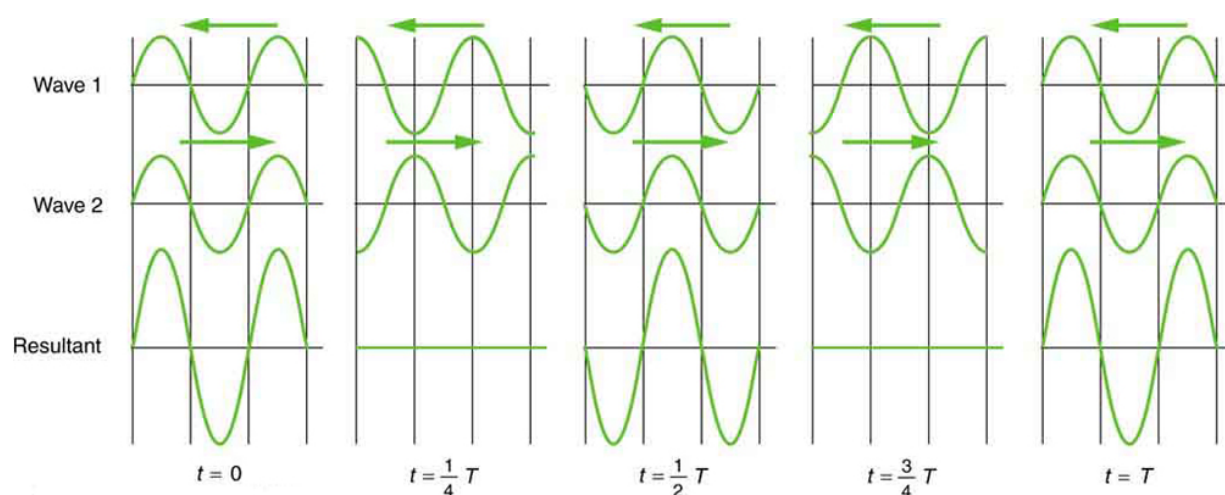


Superposition of non-identical waves exhibits both constructive and destructive interference.

Standing Waves

Sometimes waves do not seem to move; rather, they just vibrate in place. Unmoving waves can be seen on the surface of a glass of milk in a refrigerator, for example. Vibrations from the refrigerator motor create waves on the milk that oscillate up and down but do not seem to move across the surface. These waves are formed by the superposition of two or more moving waves, such as illustrated in [\[link\]](#) for two identical waves moving in opposite directions. The waves move through each other with their disturbances adding as they go by. If the two waves have the same amplitude and wavelength, then they alternate between constructive and destructive interference. The resultant looks like a wave standing in place and, thus, is called a **standing wave**. Waves on the glass of milk are one example of standing waves. There are other standing waves, such as on guitar strings and in organ pipes. With the glass of milk, the two waves that produce standing waves may come from reflections from the side of the glass.

A closer look at earthquakes provides evidence for conditions appropriate for resonance, standing waves, and constructive and destructive interference. A building may be vibrated for several seconds with a driving frequency matching that of the natural frequency of vibration of the building—producing a resonance resulting in one building collapsing while neighboring buildings do not. Often buildings of a certain height are devastated while other taller buildings remain intact. The building height matches the condition for setting up a standing wave for that particular height. As the earthquake waves travel along the surface of Earth and reflect off denser rocks, constructive interference occurs at certain points. Often areas closer to the epicenter are not damaged while areas farther away are damaged.

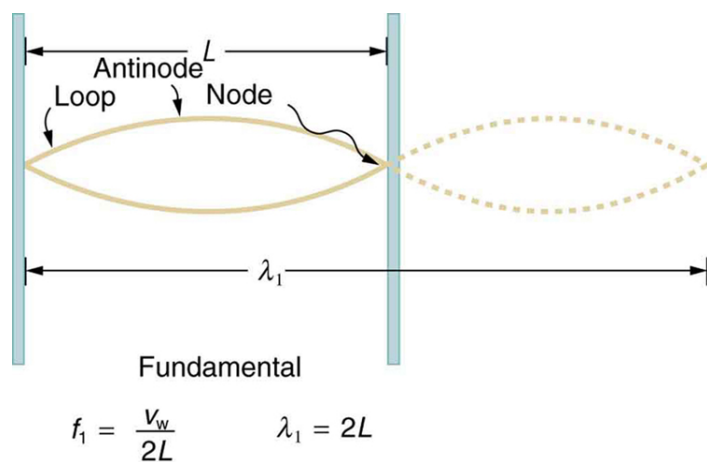


Standing wave created by the superposition of two identical waves moving in opposite directions. The oscillations are at fixed locations in space and result from alternately constructive and destructive interference.

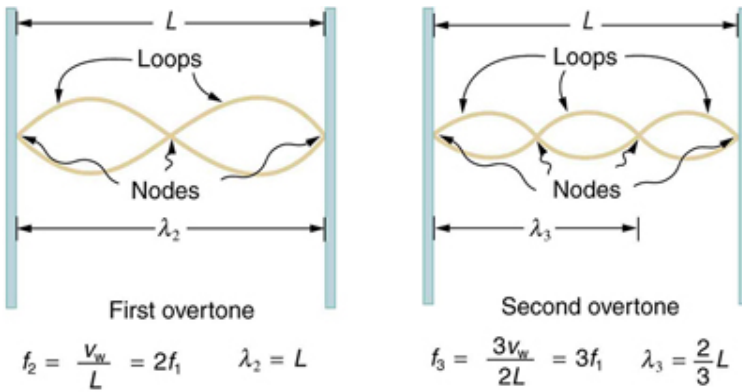
Standing waves are also found on the strings of musical instruments and are due to reflections of waves from the ends of the string. [\[link\]](#) and [\[link\]](#) show three standing waves that can be created on a string that is fixed at both ends. **Nodes** are the points where the string does not move; more

generally, nodes are where the wave disturbance is zero in a standing wave. The fixed ends of strings must be nodes, too, because the string cannot move there. The word **antinode** is used to denote the location of maximum amplitude in standing waves. Standing waves on strings have a frequency that is related to the propagation speed v_w of the disturbance on the string. The wavelength λ is determined by the distance between the points where the string is fixed in place.

The lowest frequency, called the **fundamental frequency**, is thus for the longest wavelength, which is seen to be $\lambda_1 = 2L$. Therefore, the fundamental frequency is $f_1 = v_w/\lambda_1 = v_w/2L$. In this case, the **overtones** or harmonics are multiples of the fundamental frequency. As seen in [\[link\]](#), the first harmonic can easily be calculated since $\lambda_2 = L$. Thus, $f_2 = v_w/\lambda_2 = v_w/2L = 2f_1$. Similarly, $f_3 = 3f_1$, and so on. All of these frequencies can be changed by adjusting the tension in the string. The greater the tension, the greater v_w is and the higher the frequencies. This observation is familiar to anyone who has ever observed a string instrument being tuned. We will see in later chapters that standing waves are crucial to many resonance phenomena, such as in sounding boxes on string instruments.



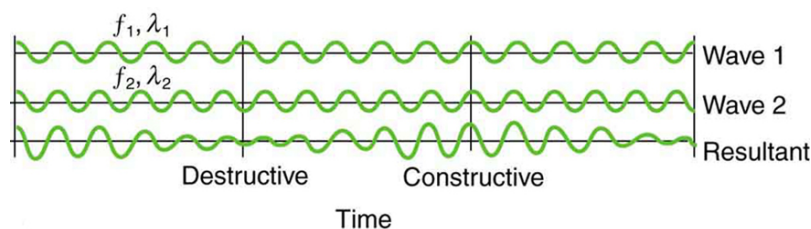
The figure shows a string oscillating at its fundamental frequency.



First and second harmonic frequencies are shown.

Beats

Striking two adjacent keys on a piano produces a warbling combination usually considered to be unpleasant. The superposition of two waves of similar but not identical frequencies is the culprit. Another example is often noticeable in jet aircraft, particularly the two-engine variety, while taxiing. The combined sound of the engines goes up and down in loudness. This varying loudness happens because the sound waves have similar but not identical frequencies. The discordant warbling of the piano and the fluctuating loudness of the jet engine noise are both due to alternately constructive and destructive interference as the two waves go in and out of phase. [\[link\]](#) illustrates this graphically.



Beats are produced by the superposition of two waves of slightly different frequencies but identical amplitudes. The waves alternate in time between constructive interference and destructive interference, giving the resulting wave a time-varying amplitude.

The wave resulting from the superposition of two similar-frequency waves has a frequency that is the average of the two. This wave fluctuates in amplitude, or *beats*, with a frequency called the **beat frequency**. We can determine the beat frequency by adding two waves together mathematically. Note that a wave can be represented at one point in space as

Equation:

$$x = X \cos\left(\frac{2\pi t}{T}\right) = X \cos(2\pi ft),$$

where $f = 1/T$ is the frequency of the wave. Adding two waves that have different frequencies but identical amplitudes produces a resultant

Equation:

$$x = x_1 + x_2.$$

More specifically,

Equation:

$$x = X \cos(2\pi f_1 t) + X \cos(2\pi f_2 t).$$

Using a trigonometric identity, it can be shown that

Equation:

$$x = 2X \cos(\pi f_B t) \cos(2\pi f_{\text{ave}} t),$$

where

Equation:

$$f_B = |f_1 - f_2|$$

is the beat frequency, and f_{ave} is the average of f_1 and f_2 . These results mean that the resultant wave has twice the amplitude and the average frequency of the two superimposed waves, but it also fluctuates in overall amplitude at the beat frequency f_B . The first cosine term in the expression effectively causes the amplitude to go up and down. The second cosine term is the wave with frequency f_{ave} . This result is valid for all types of waves. However, if it is a sound wave, providing the two frequencies are similar, then what we hear is an average frequency that gets louder and softer (or warbles) at the beat frequency.

Note:

Making Career Connections

Piano tuners use beats routinely in their work. When comparing a note with a tuning fork, they listen for beats and adjust the string until the beats go away (to zero frequency). For example, if the tuning fork has a 256 Hz frequency and two beats per second are heard, then the other frequency is either 254 or 258 Hz. Most keys hit multiple strings, and these strings are actually adjusted until they have nearly the same frequency and give a slow beat for richness. Twelve-string guitars and mandolins are also tuned using beats.

While beats may sometimes be annoying in audible sounds, we will find that beats have many applications. Observing beats is a very useful way to compare similar frequencies. There are applications of beats as apparently disparate as in ultrasonic imaging and radar speed traps.

Exercise:

Check Your Understanding

Problem:

Imagine you are holding one end of a jump rope, and your friend holds the other. If your friend holds her end still, you can move your end up and down, creating a transverse wave. If your friend then begins to move her end up and down, generating a wave in the opposite direction, what resultant wave forms would you expect to see in the jump rope?

Solution:

The rope would alternate between having waves with amplitudes two times the original amplitude and reaching equilibrium with no amplitude at all. The wavelengths will result in both constructive and destructive interference

Exercise:

Check Your Understanding

Problem: Define nodes and antinodes.

Solution:

Nodes are areas of wave interference where there is no motion. Antinodes are areas of wave interference where the motion is at its maximum point.

Exercise:

Check Your Understanding

Problem:

You hook up a stereo system. When you test the system, you notice that in one corner of the room, the sounds seem dull. In another area, the sounds seem excessively loud. Describe how the sound moving about the room could result in these effects.

Solution:

With multiple speakers putting out sounds into the room, and these sounds bouncing off walls, there is bound to be some wave interference. In the dull areas, the interference is probably mostly destructive. In the louder areas, the interference is probably mostly constructive.

Note:

PhET Explorations: Wave Interference

Make waves with a dripping faucet, audio speaker, or laser! Add a second source or a pair of slits to create an interference pattern.

[Wave](#)
[Interferenc](#)
[e](#)

Section Summary

- Superposition is the combination of two waves at the same location.
- Constructive interference occurs when two identical waves are superimposed in phase.

- Destructive interference occurs when two identical waves are superimposed exactly out of phase.
- A standing wave is one in which two waves superimpose to produce a wave that varies in amplitude but does not propagate.
- Nodes are points of no motion in standing waves.
- An antinode is the location of maximum amplitude of a standing wave.
- Waves on a string are resonant standing waves with a fundamental frequency and can occur at higher multiples of the fundamental, called overtones or harmonics.
- Beats occur when waves of similar frequencies f_1 and f_2 are superimposed. The resulting amplitude oscillates with a beat frequency given by
Equation:

$$f_B = |f_1 - f_2|.$$

Conceptual Questions

Exercise:

Problem:

Speakers in stereo systems have two color-coded terminals to indicate how to hook up the wires. If the wires are reversed, the speaker moves in a direction opposite that of a properly connected speaker. Explain why it is important to have both speakers connected the same way.

Problems & Exercises

Exercise:

Problem:

A car has two horns, one emitting a frequency of 199 Hz and the other emitting a frequency of 203 Hz. What beat frequency do they produce?

Solution:

$$f = 4 \text{ Hz}$$

Exercise:**Problem:**

The middle-C hammer of a piano hits two strings, producing beats of 1.50 Hz. One of the strings is tuned to 260.00 Hz. What frequencies could the other string have?

Exercise:**Problem:**

Two tuning forks having frequencies of 460 and 464 Hz are struck simultaneously. What average frequency will you hear, and what will the beat frequency be?

Solution:

462 Hz,

4 Hz

Exercise:**Problem:**

Twin jet engines on an airplane are producing an average sound frequency of 4100 Hz with a beat frequency of 0.500 Hz. What are their individual frequencies?

Exercise:**Problem:**

A wave traveling on a Slinky® that is stretched to 4 m takes 2.4 s to travel the length of the Slinky and back again. (a) What is the speed of the wave? (b) Using the same Slinky stretched to the same length, a standing wave is created which consists of three antinodes and four nodes. At what frequency must the Slinky be oscillating?

Solution:

(a) 3.33 m/s

(b) 1.25 Hz

Exercise:**Problem:**

Three adjacent keys on a piano (F, F-sharp, and G) are struck simultaneously, producing frequencies of 349, 370, and 392 Hz. What beat frequencies are produced by this discordant combination?

Glossary

antinode

the location of maximum amplitude in standing waves

beat frequency

the frequency of the amplitude fluctuations of a wave

constructive interference

when two waves arrive at the same point exactly in phase; that is, the crests of the two waves are precisely aligned, as are the troughs

destructive interference

when two identical waves arrive at the same point exactly out of phase; that is, precisely aligned crest to trough

fundamental frequency

the lowest frequency of a periodic waveform

nodes

the points where the string does not move; more generally, nodes are where the wave disturbance is zero in a standing wave

overtones

multiples of the fundamental frequency of a sound

superposition

the phenomenon that occurs when two or more waves arrive at the same point

Energy in Waves: Intensity

- Calculate the intensity and the power of rays and waves.



The destructive effect of an earthquake is palpable evidence of the energy carried in these waves. The Richter scale rating of earthquakes is related to both their amplitude and the energy they carry.

(credit: Petty Officer 2nd Class Candice Villarreal, U.S. Navy)

All waves carry energy. The energy of some waves can be directly observed. Earthquakes can shake whole cities to the ground, performing the work of thousands of wrecking balls.

Loud sounds pulverize nerve cells in the inner ear, causing permanent hearing loss. Ultrasound is used for deep-heat treatment of muscle strains. A laser beam can burn away a malignancy. Water waves chew up beaches.

The amount of energy in a wave is related to its amplitude. Large-amplitude earthquakes produce large ground displacements. Loud sounds have higher pressure amplitudes and come from larger-amplitude source vibrations than

soft sounds. Large ocean breakers churn up the shore more than small ones. More quantitatively, a wave is a displacement that is resisted by a restoring force. The larger the displacement x , the larger the force $F = kx$ needed to create it. Because work W is related to force multiplied by distance (Fx) and energy is put into the wave by the work done to create it, the energy in a wave is related to amplitude. In fact, a wave's energy is directly proportional to its amplitude squared because

Equation:

$$W \propto Fx = kx^2.$$

The energy effects of a wave depend on time as well as amplitude. For example, the longer deep-heat ultrasound is applied, the more energy it transfers. Waves can also be concentrated or spread out. Sunlight, for example, can be focused to burn wood. Earthquakes spread out, so they do less damage the farther they get from the source. In both cases, changing the area the waves cover has important effects. All these pertinent factors are included in the definition of **intensity** I as power per unit area:

Equation:

$$I = \frac{P}{A}$$

where P is the power carried by the wave through area A . The definition of intensity is valid for any energy in transit, including that carried by waves. The SI unit for intensity is watts per square meter (W/m^2). For example, infrared and visible energy from the Sun impinge on Earth at an intensity of $1300 \text{ W}/\text{m}^2$ just above the atmosphere. There are other intensity-related units in use, too. The most common is the decibel. For example, a 90 decibel sound level corresponds to an intensity of $10^{-3} \text{ W}/\text{m}^2$. (This quantity is not much power per unit area considering that 90 decibels is a relatively high sound level. Decibels will be discussed in some detail in a later chapter.

Example:**Calculating intensity and power: How much energy is in a ray of sunlight?**

The average intensity of sunlight on Earth's surface is about 700 W/m^2 .

(a) Calculate the amount of energy that falls on a solar collector having an area of 0.500 m^2 in 4.00 h .

(b) What intensity would such sunlight have if concentrated by a magnifying glass onto an area 200 times smaller than its own?

Strategy a

Because power is energy per unit time or $P = \frac{E}{t}$, the definition of intensity can be written as $I = \frac{P}{A} = \frac{E/t}{A}$, and this equation can be solved for E with the given information.

Solution a

1. Begin with the equation that states the definition of intensity:

Equation:

$$I = \frac{P}{A}.$$

2. Replace P with its equivalent E/t :

Equation:

$$I = \frac{E/t}{A}.$$

3. Solve for E :

Equation:

$$E = IAt.$$

4. Substitute known values into the equation:

Equation:

$$E = (700 \text{ W/m}^2)(0.500 \text{ m}^2)[(4.00 \text{ h})(3600 \text{ s/h})].$$

5. Calculate to find E and convert units:

Equation:

$$5.04 \times 10^6 \text{ J},$$

Discussion a

The energy falling on the solar collector in 4 h in part is enough to be useful—for example, for heating a significant amount of water.

Strategy b

Taking a ratio of new intensity to old intensity and using primes for the new quantities, we will find that it depends on the ratio of the areas. All other quantities will cancel.

Solution b

1. Take the ratio of intensities, which yields:

Equation:

$$\frac{I'}{I} = \frac{P'/A'}{P/A} = \frac{A}{A'} \left(\text{The powers cancel because } P' = P \right).$$

2. Identify the knowns:

Equation:

$$A = 200A',$$

Equation:

$$\frac{I'}{I} = 200.$$

3. Substitute known quantities:

Equation:

$$I' = 200I = 200(700 \text{ W/m}^2).$$

4. Calculate to find I' :

Equation:

$$I' = 1.40 \times 10^5 \text{ W/m}^2.$$

Discussion b

Decreasing the area increases the intensity considerably. The intensity of the concentrated sunlight could even start a fire.

Example:**Determine the combined intensity of two waves: Perfect constructive interference**

If two identical waves, each having an intensity of 1.00 W/m^2 , interfere perfectly constructively, what is the intensity of the resulting wave?

Strategy

We know from [Superposition and Interference](#) that when two identical waves, which have equal amplitudes X , interfere perfectly constructively, the resulting wave has an amplitude of $2X$. Because a wave's intensity is proportional to amplitude squared, the intensity of the resulting wave is four times as great as in the individual waves.

Solution

1. Recall that intensity is proportional to amplitude squared.
2. Calculate the new amplitude:

Equation:

$$I' \propto (X')^2 = (2X)^2 = 4X^2.$$

3. Recall that the intensity of the old amplitude was:

Equation:

$$I \propto X^2.$$

4. Take the ratio of new intensity to the old intensity. This gives:

Equation:

$$\frac{I'}{I} = 4.$$

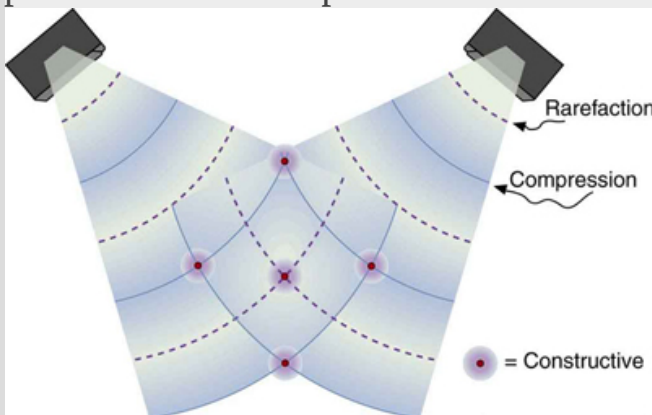
5. Calculate to find I' :

Equation:

$$I' = 4I = 4.00 \text{ W/m}^2.$$

Discussion

The intensity goes up by a factor of 4 when the amplitude doubles. This answer is a little disquieting. The two individual waves each have intensities of 1.00 W/m^2 , yet their sum has an intensity of 4.00 W/m^2 , which may appear to violate conservation of energy. This violation, of course, cannot happen. What does happen is intriguing. The area over which the intensity is 4.00 W/m^2 is much less than the area covered by the two waves before they interfered. There are other areas where the intensity is zero. The addition of waves is not as simple as our first look in [Superposition and Interference](#) suggested. We actually get a pattern of both constructive interference and destructive interference whenever two waves are added. For example, if we have two stereo speakers putting out 1.00 W/m^2 each, there will be places in the room where the intensity is 4.00 W/m^2 , other places where the intensity is zero, and others in between. [\[link\]](#) shows what this interference might look like. We will pursue interference patterns elsewhere in this text.



These stereo speakers produce both constructive interference and destructive interference in the room, a property common to the

superposition of all types of waves.
The shading is proportional to
intensity.

Exercise:

Check Your Understanding

Problem:

Which measurement of a wave is most important when determining the wave's intensity?

Solution:

Amplitude, because a wave's energy is directly proportional to its amplitude squared.

Section Summary

Intensity is defined to be the power per unit area:

$$I = \frac{P}{A} \text{ and has units of } \text{W}/\text{m}^2.$$

Conceptual Questions

Exercise:

Problem:

Two identical waves undergo pure constructive interference. Is the resultant intensity twice that of the individual waves? Explain your answer.

Exercise:

Problem:

Circular water waves decrease in amplitude as they move away from where a rock is dropped. Explain why.

Problems & Exercises**Exercise:****Problem: Medical Application**

Ultrasound of intensity $1.50 \times 10^2 \text{ W/m}^2$ is produced by the rectangular head of a medical imaging device measuring 3.00 by 5.00 cm. What is its power output?

Solution:

0.225 W

Exercise:**Problem:**

The low-frequency speaker of a stereo set has a surface area of 0.05 m^2 and produces 1W of acoustical power. What is the intensity at the speaker? If the speaker projects sound uniformly in all directions, at what distance from the speaker is the intensity 0.1 W/m^2 ?

Exercise:**Problem:**

To increase intensity of a wave by a factor of 50, by what factor should the amplitude be increased?

Solution:

7.07

Exercise:**Problem: Engineering Application**

A device called an insolation meter is used to measure the intensity of sunlight has an area of 100 cm^2 and registers 6.50 W . What is the intensity in W/m^2 ?

Exercise:**Problem: Astronomy Application**

Energy from the Sun arrives at the top of the Earth's atmosphere with an intensity of 1.30 kW/m^2 . How long does it take for $1.8 \times 10^9 \text{ J}$ to arrive on an area of 1.00 m^2 ?

Solution:

16.0 d

Exercise:**Problem:**

Suppose you have a device that extracts energy from ocean breakers in direct proportion to their intensity. If the device produces 10.0 kW of power on a day when the breakers are 1.20 m high, how much will it produce when they are 0.600 m high?

Solution:

2.50 kW

Exercise:**Problem: Engineering Application**

(a) A photovoltaic array of (solar cells) is 10.0% efficient in gathering solar energy and converting it to electricity. If the average intensity of

sunlight on one day is 700 W/m^2 , what area should your array have to gather energy at the rate of 100 W ? (b) What is the maximum cost of the array if it must pay for itself in two years of operation averaging 10.0 hours per day? Assume that it earns money at the rate of 9.00 ¢ per kilowatt-hour.

Exercise:

Problem:

A microphone receiving a pure sound tone feeds an oscilloscope, producing a wave on its screen. If the sound intensity is originally $2.00 \times 10^{-5} \text{ W/m}^2$, but is turned up until the amplitude increases by 30.0% , what is the new intensity?

Solution:

$$3.38 \times 10^{-5} \text{ W/m}^2$$

Exercise:

Problem: Medical Application

(a) What is the intensity in W/m^2 of a laser beam used to burn away cancerous tissue that, when 90.0% absorbed, puts 500 J of energy into a circular spot 2.00 mm in diameter in 4.00 s ? (b) Discuss how this intensity compares to the average intensity of sunlight (about 700 W/m^2) and the implications that would have if the laser beam entered your eye. Note how your answer depends on the time duration of the exposure.

Glossary

intensity

power per unit area

Introduction to the Physics of Hearing

class="introduction"

This tree fell
some time
ago. When it
fell, atoms in
the air were
disturbed.
Physicists
would call
this
disturbance
sound
whether
someone was
around to
hear it or not.
(credit: B.A.
Bowen
Photography
)



If a tree falls in the forest and no one is there to hear it, does it make a sound? The answer to this old philosophical question depends on how you define sound. If sound only exists when someone is around to perceive it, then there was no sound. However, if we define sound in terms of physics; that is, a disturbance of the atoms in matter transmitted from its origin outward (in other words, a wave), then there *was* a sound, even if nobody was around to hear it.

Such a wave is the physical phenomenon we call *sound*. Its perception is hearing. Both the physical phenomenon and its perception are interesting and will be considered in this text. We shall explore both sound and hearing; they are related, but are not the same thing. We will also explore the many practical uses of sound waves, such as in medical imaging.

Sound

- Define sound and hearing.
- Describe sound as a longitudinal wave.



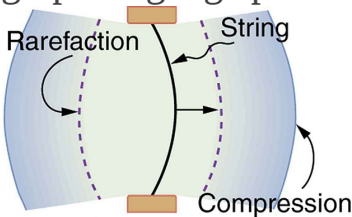
This glass has been shattered by a high-intensity sound wave of the same frequency as the resonant frequency of the glass. While the sound is not visible, the effects of the sound prove its existence.

(credit: ||read||,
Flickr)

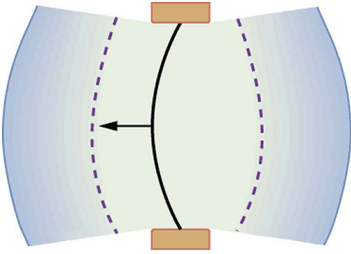
Sound can be used as a familiar illustration of waves. Because hearing is one of our most important senses, it is interesting to see how the physical properties of sound correspond to our perceptions of it. **Hearing** is the perception of sound, just as vision is the perception of visible light. But sound has important applications beyond hearing. Ultrasound, for example, is not heard but can be employed to form medical images and is also used in treatment.

The physical phenomenon of **sound** is defined to be a disturbance of matter that is transmitted from its source outward. Sound is a wave. On the atomic scale, it is a disturbance of atoms that is far more ordered than their thermal motions. In many instances, sound is a periodic wave, and the atoms undergo simple harmonic motion. In this text, we shall explore such periodic sound waves.

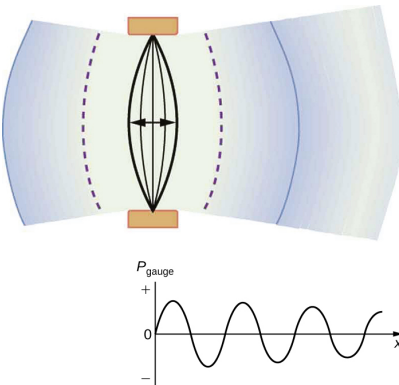
A vibrating string produces a sound wave as illustrated in [\[link\]](#), [\[link\]](#), and [\[link\]](#). As the string oscillates back and forth, it transfers energy to the air, mostly as thermal energy created by turbulence. But a small part of the string's energy goes into compressing and expanding the surrounding air, creating slightly higher and lower local pressures. These compressions (high pressure regions) and rarefactions (low pressure regions) move out as longitudinal pressure waves having the same frequency as the string—they are the disturbance that is a sound wave. (Sound waves in air and most fluids are longitudinal, because fluids have almost no shear strength. In solids, sound waves can be both transverse and longitudinal.) [\[link\]](#) shows a graph of gauge pressure versus distance from the vibrating string.



A vibrating
string moving to
the right
compresses the
air in front of it
and expands the
air behind it.



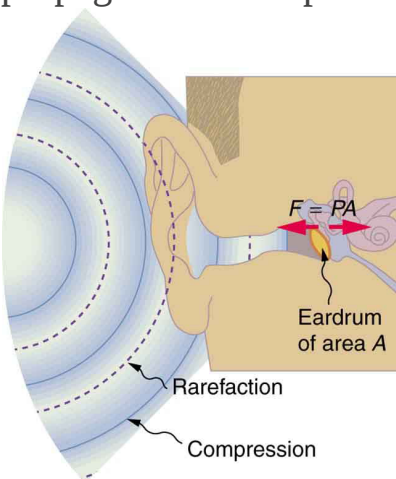
As the string moves to the left, it creates another compression and rarefaction as the ones on the right move away from the string.



After many vibrations, there are a series of compressions and rarefactions moving out from the string as a sound wave. The graph shows gauge pressure versus

distance from the source. Pressures vary only slightly from atmospheric for ordinary sounds.

The amplitude of a sound wave decreases with distance from its source, because the energy of the wave is spread over a larger and larger area. But it is also absorbed by objects, such as the eardrum in [\[link\]](#), and converted to thermal energy by the viscosity of air. In addition, during each compression a little heat transfers to the air and during each rarefaction even less heat transfers from the air, so that the heat transfer reduces the organized disturbance into random thermal motions. (These processes can be viewed as a manifestation of the second law of thermodynamics presented in [Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency](#).) Whether the heat transfer from compression to rarefaction is significant depends on how far apart they are—that is, it depends on wavelength. Wavelength, frequency, amplitude, and speed of propagation are important for sound, as they are for all waves.



Sound wave
compressions and
rarefactions travel
up the ear canal and

force the eardrum to vibrate. There is a net force on the eardrum, since the sound wave pressures differ from the atmospheric pressure found behind the eardrum. A complicated mechanism converts the vibrations to nerve impulses, which are perceived by the person.

Note:

PhET Explorations: Wave Interference

WMake waves with a dripping faucet, audio speaker, or laser! Add a second source or a pair of slits to create an interference pattern.

<https://archive.cnx.org/specials/2fe7ad15-b00e-4402-b068-ff503985a18f/wave-interference/>

Section Summary

- Sound is a disturbance of matter that is transmitted from its source outward.
- Sound is one type of wave.

- Hearing is the perception of sound.

Glossary

sound

a disturbance of matter that is transmitted from its source outward

hearing

the perception of sound

Speed of Sound, Frequency, and Wavelength

- Define pitch.
- Describe the relationship between the speed of sound, its frequency, and its wavelength.
- Describe the effects on the speed of sound as it travels through various media.
- Describe the effects of temperature on the speed of sound.



When a firework explodes, the light energy is perceived before the sound energy. Sound travels more slowly than light does.
(credit: Dominic Alves, Flickr)

Sound, like all waves, travels at a certain speed and has the properties of frequency and wavelength. You can observe direct evidence of the speed of sound while watching a fireworks display. The flash of an explosion is seen well before its sound is heard, implying both that sound travels at a finite speed and that it is much slower than light. You can also directly sense the frequency of a sound. Perception of frequency is called **pitch**. The wavelength of sound is not directly sensed, but indirect evidence is found in the correlation of the size of musical instruments with their pitch. Small

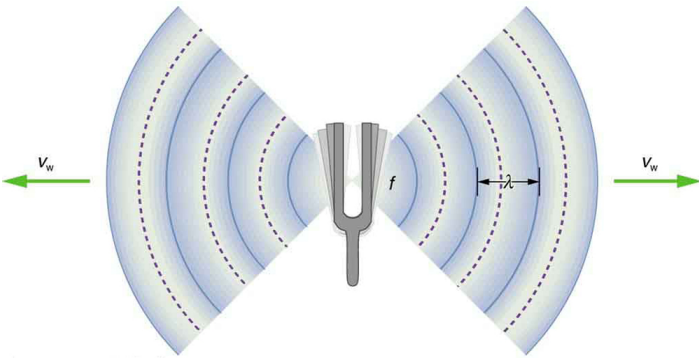
instruments, such as a piccolo, typically make high-pitch sounds, while large instruments, such as a tuba, typically make low-pitch sounds. High pitch means small wavelength, and the size of a musical instrument is directly related to the wavelengths of sound it produces. So a small instrument creates short-wavelength sounds. Similar arguments hold that a large instrument creates long-wavelength sounds.

The relationship of the speed of sound, its frequency, and wavelength is the same as for all waves:

Equation:

$$v_w = f\lambda,$$

where v_w is the speed of sound, f is its frequency, and λ is its wavelength. The wavelength of a sound is the distance between adjacent identical parts of a wave—for example, between adjacent compressions as illustrated in [\[link\]](#). The frequency is the same as that of the source and is the number of waves that pass a point per unit time.



A sound wave emanates from a source vibrating at a frequency f , propagates at v_w , and has a wavelength λ .

[\[link\]](#) makes it apparent that the speed of sound varies greatly in different media. The speed of sound in a medium is determined by a combination of the medium's rigidity (or compressibility in gases) and its density. The

more rigid (or less compressible) the medium, the faster the speed of sound. This observation is analogous to the fact that the frequency of a simple harmonic motion is directly proportional to the stiffness of the oscillating object. The greater the density of a medium, the slower the speed of sound. This observation is analogous to the fact that the frequency of a simple harmonic motion is inversely proportional to the mass of the oscillating object. The speed of sound in air is low, because air is compressible. Because liquids and solids are relatively rigid and very difficult to compress, the speed of sound in such media is generally greater than in gases.

Medium	$v_w(\text{m/s})$
<i>Gases at 0°C</i>	
Air	331
Carbon dioxide	259
Oxygen	316
Helium	965
Hydrogen	1290
<i>Liquids at 20°C</i>	
Ethanol	1160
Mercury	1450
Water, fresh	1480

Medium	$v_w(\text{m/s})$
Sea water	1540
Human tissue	1540
<i>Solids (longitudinal or bulk)</i>	
Vulcanized rubber	54
Polyethylene	920
Marble	3810
Glass, Pyrex	5640
Lead	1960
Aluminum	5120
Steel	5960

Speed of Sound in Various Media

Earthquakes, essentially sound waves in Earth's crust, are an interesting example of how the speed of sound depends on the rigidity of the medium. Earthquakes have both longitudinal and transverse components, and these travel at different speeds. The bulk modulus of granite is greater than its shear modulus. For that reason, the speed of longitudinal or pressure waves (P-waves) in earthquakes in granite is significantly higher than the speed of transverse or shear waves (S-waves). Both components of earthquakes travel slower in less rigid material, such as sediments. P-waves have speeds of 4 to 7 km/s, and S-waves correspondingly range in speed from 2 to 5 km/s, both being faster in more rigid material. The P-wave gets progressively farther ahead of the S-wave as they travel through Earth's crust. The time between the P- and S-waves is routinely used to determine the distance to their source, the epicenter of the earthquake.

The speed of sound is affected by temperature in a given medium. For air at sea level, the speed of sound is given by

Equation:

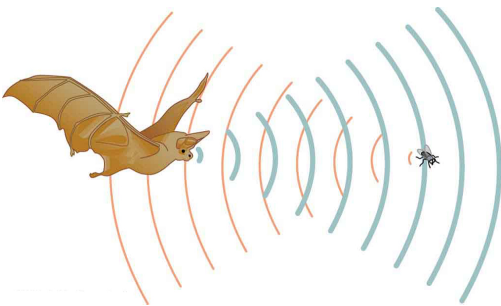
$$v_w = (331 \text{ m/s}) \sqrt{\frac{T}{273 \text{ K}}},$$

where the temperature (denoted as T) is in units of kelvin. The speed of sound in gases is related to the average speed of particles in the gas, v_{rms} , and that

Equation:

$$v_{\text{rms}} = \sqrt{\frac{3 kT}{m}},$$

where k is the Boltzmann constant ($1.38 \times 10^{-23} \text{ J/K}$) and m is the mass of each (identical) particle in the gas. So, it is reasonable that the speed of sound in air and other gases should depend on the square root of temperature. While not negligible, this is not a strong dependence. At 0°C , the speed of sound is 331 m/s, whereas at 20.0°C it is 343 m/s, less than a 4% increase. [\[link\]](#) shows a use of the speed of sound by a bat to sense distances. Echoes are also used in medical imaging.



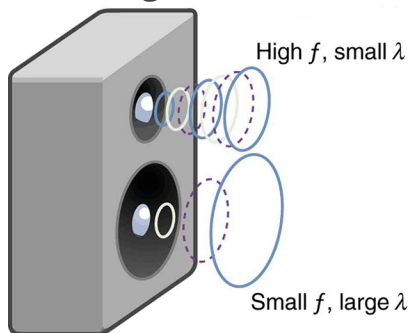
A bat uses sound echoes to find its way about and to catch prey. The time for the echo to return is directly proportional to the distance.

One of the more important properties of sound is that its speed is nearly independent of frequency. This independence is certainly true in open air for sounds in the audible range of 20 to 20,000 Hz. If this independence were not true, you would certainly notice it for music played by a marching band in a football stadium, for example. Suppose that high-frequency sounds traveled faster—then the farther you were from the band, the more the sound from the low-pitch instruments would lag that from the high-pitch ones. But the music from all instruments arrives in cadence independent of distance, and so all frequencies must travel at nearly the same speed. Recall that

Equation:

$$v_w = f\lambda.$$

In a given medium under fixed conditions, v_w is constant, so that there is a relationship between f and λ ; the higher the frequency, the smaller the wavelength. See [\[link\]](#) and consider the following example.



Because they travel
at the same speed
in a given medium,
low-frequency
sounds must have a
greater wavelength
than high-
frequency sounds.

Here, the lower-frequency sounds are emitted by the large speaker, called a woofer, while the higher-frequency sounds are emitted by the small speaker, called a tweeter.

Example:

Calculating Wavelengths: What Are the Wavelengths of Audible Sounds?

Calculate the wavelengths of sounds at the extremes of the audible range, 20 and 20,000 Hz, in 30.0°C air. (Assume that the frequency values are accurate to two significant figures.)

Strategy

To find wavelength from frequency, we can use $v_w = f\lambda$.

Solution

1. Identify knowns. The value for v_w , is given by

Equation:

$$v_w = (331 \text{ m/s}) \sqrt{\frac{T}{273 \text{ K}}}.$$

2. Convert the temperature into kelvin and then enter the temperature into the equation

Equation:

$$v_w = (331 \text{ m/s}) \sqrt{\frac{303 \text{ K}}{273 \text{ K}}} = 348.7 \text{ m/s}.$$

3. Solve the relationship between speed and wavelength for λ :

Equation:

$$\lambda = \frac{v_w}{f}.$$

4. Enter the speed and the minimum frequency to give the maximum wavelength:

Equation:

$$\lambda_{\max} = \frac{348.7 \text{ m/s}}{20 \text{ Hz}} = 17 \text{ m}.$$

5. Enter the speed and the maximum frequency to give the minimum wavelength:

Equation:

$$\lambda_{\min} = \frac{348.7 \text{ m/s}}{20,000 \text{ Hz}} = 0.017 \text{ m} = 1.7 \text{ cm}.$$

Discussion

Because the product of f multiplied by λ equals a constant, the smaller f is, the larger λ must be, and vice versa.

The speed of sound can change when sound travels from one medium to another. However, the frequency usually remains the same because it is like a driven oscillation and has the frequency of the original source. If v_w changes and f remains the same, then the wavelength λ must change. That is, because $v_w = f\lambda$, the higher the speed of a sound, the greater its wavelength for a given frequency.

Note:

Making Connections: Take-Home Investigation—Voice as a Sound Wave

Suspend a sheet of paper so that the top edge of the paper is fixed and the bottom edge is free to move. You could tape the top edge of the paper to the edge of a table. Gently blow near the edge of the bottom of the sheet and note how the sheet moves. Speak softly and then louder such that the sounds hit the edge of the bottom of the paper, and note how the sheet moves. Explain the effects.

Exercise:

Check Your Understanding

Problem:

Imagine you observe two fireworks explode. You hear the explosion of one as soon as you see it. However, you see the other firework for several milliseconds before you hear the explosion. Explain why this is so.

Solution:

Sound and light both travel at definite speeds. The speed of sound is slower than the speed of light. The first firework is probably very close by, so the speed difference is not noticeable. The second firework is farther away, so the light arrives at your eyes noticeably sooner than the sound wave arrives at your ears.

Exercise:

Check Your Understanding

Problem:

You observe two musical instruments that you cannot identify. One plays high-pitch sounds and the other plays low-pitch sounds. How could you determine which is which without hearing either of them play?

Solution:

Compare their sizes. High-pitch instruments are generally smaller than low-pitch instruments because they generate a smaller wavelength.

Section Summary

The relationship of the speed of sound v_w , its frequency f , and its wavelength λ is given by

Equation:

$$v_w = f\lambda,$$

which is the same relationship given for all waves.

In air, the speed of sound is related to air temperature T by

Equation:

$$v_w = (331 \text{ m/s}) \sqrt{\frac{T}{273 \text{ K}}}.$$

v_w is the same for all frequencies and wavelengths.

Conceptual Questions

Exercise:

Problem:

How do sound vibrations of atoms differ from thermal motion?

Exercise:

Problem:

When sound passes from one medium to another where its propagation speed is different, does its frequency or wavelength change? Explain your answer briefly.

Problems & Exercises

Exercise:

Problem:

When poked by a spear, an operatic soprano lets out a 1200-Hz shriek. What is its wavelength if the speed of sound is 345 m/s?

Solution:

0.288 m

Exercise:

Problem:

What frequency sound has a 0.10-m wavelength when the speed of sound is 340 m/s?

Exercise:

Problem:

Calculate the speed of sound on a day when a 1500 Hz frequency has a wavelength of 0.221 m.

Solution:

332 m/s

Exercise:

Problem:

(a) What is the speed of sound in a medium where a 100-kHz frequency produces a 5.96-cm wavelength? (b) Which substance in [\[link\]](#) is this likely to be?

Exercise:

Problem:

Show that the speed of sound in 20.0°C air is 343 m/s, as claimed in the text.

Solution:**Equation:**

$$\begin{aligned}v_w &= (331 \text{ m/s}) \sqrt{\frac{T}{273 \text{ K}}} = (331 \text{ m/s}) \sqrt{\frac{293 \text{ K}}{273 \text{ K}}} \\&= 343 \text{ m/s}\end{aligned}$$

Exercise:**Problem:**

Air temperature in the Sahara Desert can reach 56.0°C (about 134°F). What is the speed of sound in air at that temperature?

Exercise:**Problem:**

Dolphins make sounds in air and water. What is the ratio of the wavelength of a sound in air to its wavelength in seawater? Assume air temperature is 20.0°C.

Solution:

0.223

Exercise:**Problem:**

A sonar echo returns to a submarine 1.20 s after being emitted. What is the distance to the object creating the echo? (Assume that the submarine is in the ocean, not in fresh water.)

Exercise:

Problem:

(a) If a submarine's sonar can measure echo times with a precision of 0.0100 s, what is the smallest difference in distances it can detect?

(Assume that the submarine is in the ocean, not in fresh water.)

(b) Discuss the limits this time resolution imposes on the ability of the sonar system to detect the size and shape of the object creating the echo.

Solution:

(a) 7.70 m

(b) This means that sonar is good for spotting and locating large objects, but it isn't able to resolve smaller objects, or detect the detailed shapes of objects. Objects like ships or large pieces of airplanes can be found by sonar, while smaller pieces must be found by other means.

Exercise:**Problem:**

A physicist at a fireworks display times the lag between seeing an explosion and hearing its sound, and finds it to be 0.400 s. (a) How far away is the explosion if air temperature is 24.0°C and if you neglect the time taken for light to reach the physicist? (b) Calculate the distance to the explosion taking the speed of light into account. Note that this distance is negligibly greater.

Exercise:

Problem:

Suppose a bat uses sound echoes to locate its insect prey, 3.00 m away. (See [link](#).) (a) Calculate the echo times for temperatures of 5.00°C and 35.0°C. (b) What percent uncertainty does this cause for the bat in locating the insect? (c) Discuss the significance of this uncertainty and whether it could cause difficulties for the bat. (In practice, the bat continues to use sound as it closes in, eliminating most of any difficulties imposed by this and other effects, such as motion of the prey.)

Solution:

(a) 18.0 ms, 17.1 ms

(b) 5.00%

(c) This uncertainty could definitely cause difficulties for the bat, if it didn't continue to use sound as it closed in on its prey. A 5% uncertainty could be the difference between catching the prey around the neck or around the chest, which means that it could miss grabbing its prey.

Glossary

pitch

the perception of the frequency of a sound

Sound Intensity and Sound Level

- Define intensity, sound intensity, and sound pressure level.
- Calculate sound intensity levels in decibels (dB).



Noise on crowded roadways like this one in Delhi makes it hard to hear others unless they shout. (credit: Lingaraj G J, Flickr)

In a quiet forest, you can sometimes hear a single leaf fall to the ground. After settling into bed, you may hear your blood pulsing through your ears. But when a passing motorist has his stereo turned up, you cannot even hear what the person next to you in your car is saying. We are all very familiar with the loudness of sounds and aware that they are related to how energetically the source is vibrating. In cartoons depicting a screaming person (or an animal making a loud noise), the cartoonist often shows an open mouth with a vibrating uvula, the hanging tissue at the back of the mouth, to suggest a loud sound coming from the throat [\[link\]](#). High noise exposure is hazardous to hearing, and it is common for musicians to have hearing losses that are sufficiently severe that they interfere with the musicians' abilities to perform. The relevant physical quantity is sound intensity, a concept that is valid for all sounds whether or not they are in the audible range.

Intensity is defined to be the power per unit area carried by a wave. Power is the rate at which energy is transferred by the wave. In equation form, **intensity** I is

Equation:

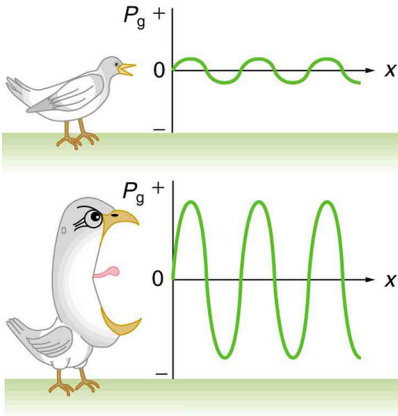
$$I = \frac{P}{A},$$

where P is the power through an area A . The SI unit for I is W/m^2 . The intensity of a sound wave is related to its amplitude squared by the following relationship:

Equation:

$$I = \frac{(\Delta p)^2}{2\rho v_w}.$$

Here Δp is the pressure variation or pressure amplitude (half the difference between the maximum and minimum pressure in the sound wave) in units of pascals (Pa) or N/m^2 . (We are using a lower case p for pressure to distinguish it from power, denoted by P above.) The energy (as kinetic energy $\frac{mv^2}{2}$) of an oscillating element of air due to a traveling sound wave is proportional to its amplitude squared. In this equation, ρ is the density of the material in which the sound wave travels, in units of kg/m^3 , and v_w is the speed of sound in the medium, in units of m/s . The pressure variation is proportional to the amplitude of the oscillation, and so I varies as $(\Delta p)^2$ ([\[link\]](#)). This relationship is consistent with the fact that the sound wave is produced by some vibration; the greater its pressure amplitude, the more the air is compressed in the sound it creates.



Graphs of the gauge pressures in two sound waves of different intensities.

The more intense sound is produced by a source that has larger-amplitude oscillations and has greater pressure maxima and minima. Because pressures are higher in the greater-intensity sound, it can exert larger forces on the objects it encounters.

Sound intensity levels are quoted in decibels (dB) much more often than sound intensities in watts per meter squared. Decibels are the unit of choice in the scientific literature as well as in the popular media. The reasons for this choice of units are related to how we perceive sounds. How our ears perceive sound can be more accurately described by the logarithm of the

intensity rather than directly to the intensity. The **sound intensity level** β in decibels of a sound having an intensity I in watts per meter squared is defined to be

Equation:

$$\beta \text{ (dB)} = 10 \log_{10}\left(\frac{I}{I_0}\right),$$

where $I_0 = 10^{-12} \text{ W/m}^2$ is a reference intensity. In particular, I_0 is the lowest or threshold intensity of sound a person with normal hearing can perceive at a frequency of 1000 Hz. Sound intensity level is not the same as intensity. Because β is defined in terms of a ratio, it is a unitless quantity telling you the *level* of the sound relative to a fixed standard (10^{-12} W/m^2 , in this case). The units of decibels (dB) are used to indicate this ratio is multiplied by 10 in its definition. The bel, upon which the decibel is based, is named for Alexander Graham Bell, the inventor of the telephone.

Sound intensity level β (dB)	Intensity $I(\text{W/m}^2)$	Example/effect
0	1×10^{-12}	Threshold of hearing at 1000 Hz
10	1×10^{-11}	Rustle of leaves
20	1×10^{-10}	Whisper at 1 m distance
30	1×10^{-9}	Quiet home

Sound intensity level β (dB)	Intensity $I(\text{W/m}^2)$	Example/effect
40	1×10^{-8}	Average home
50	1×10^{-7}	Average office, soft music
60	1×10^{-6}	Normal conversation
70	1×10^{-5}	Noisy office, busy traffic
80	1×10^{-4}	Loud radio, classroom lecture
90	1×10^{-3}	Inside a heavy truck; damage from prolonged exposure [footnote] Several government agencies and health-related professional associations recommend that 85 dB not be exceeded for 8-hour daily exposures in the absence of hearing protection.
100	1×10^{-2}	Noisy factory, siren at 30 m; damage from 8 h per day exposure
110	1×10^{-1}	Damage from 30 min per day exposure
120	1	Loud rock concert, pneumatic chipper at 2 m; threshold of pain
140	1×10^2	Jet airplane at 30 m; severe pain, damage in seconds
160	1×10^4	Bursting of eardrums

Sound Intensity Levels and Intensities

Sound Intensity Levels and Intensities

The decibel level of a sound having the threshold intensity of 10^{-12} W/m^2 is $\beta = 0 \text{ dB}$, because $\log_{10} 1 = 0$. That is, the threshold of hearing is 0 decibels. [\[link\]](#) gives levels in decibels and intensities in watts per meter squared for some familiar sounds.

One of the more striking things about the intensities in [\[link\]](#) is that the intensity in watts per meter squared is quite small for most sounds. The ear is sensitive to as little as a trillionth of a watt per meter squared—even more impressive when you realize that the area of the eardrum is only about 1 cm^2 , so that only 10^{-16} W falls on it at the threshold of hearing! Air molecules in a sound wave of this intensity vibrate over a distance of less than one molecular diameter, and the gauge pressures involved are less than 10^{-9} atm .

Another impressive feature of the sounds in [\[link\]](#) is their numerical range. Sound intensity varies by a factor of 10^{12} from threshold to a sound that causes damage in seconds. You are unaware of this tremendous range in sound intensity because how your ears respond can be described approximately as the logarithm of intensity. Thus, sound intensity levels in decibels fit your experience better than intensities in watts per meter squared. The decibel scale is also easier to relate to because most people are more accustomed to dealing with numbers such as 0, 53, or 120 than numbers such as 1.00×10^{-11} .

One more observation readily verified by examining [\[link\]](#) or using $I = \frac{(\Delta p)^2}{2\rho v_w}$ is that each factor of 10 in intensity corresponds to 10 dB. For example, a 90 dB sound compared with a 60 dB sound is 30 dB greater, or three factors of 10 (that is, 10^3 times) as intense. Another example is that if one sound is 10^7 as intense as another, it is 70 dB higher. See [\[link\]](#).

I_2/I_1	$\beta_2 - \beta_1$
2.0	3.0 dB
5.0	7.0 dB
10.0	10.0 dB

Ratios of Intensities and Corresponding Differences in Sound Intensity Levels

Example:

Calculating Sound Intensity Levels: Sound Waves

Calculate the sound intensity level in decibels for a sound wave traveling in air at 0°C and having a pressure amplitude of 0.656 Pa.

Strategy

We are given Δp , so we can calculate I using the equation

$I = (\Delta p)^2 / (2\rho v_w)^2$. Using I , we can calculate β straight from its definition in $\beta \text{ (dB)} = 10 \log_{10}(I/I_0)$.

Solution

(1) Identify knowns:

Sound travels at 331 m/s in air at 0°C.

Air has a density of 1.29 kg/m³ at atmospheric pressure and 0°C.

(2) Enter these values and the pressure amplitude into $I = (\Delta p)^2 / (2\rho v_w)$:

Equation:

$$I = \frac{(\Delta p)^2}{2\rho v_w} = \frac{(0.656 \text{ Pa})^2}{2(1.29 \text{ kg/m}^3)(331 \text{ m/s})} = 5.04 \times 10^{-4} \text{ W/m}^2.$$

(3) Enter the value for I and the known value for I_0 into

$\beta \text{ (dB)} = 10 \log_{10}(I/I_0)$. Calculate to find the sound intensity level in decibels:

Equation:

$$10 \log_{10}(5.04 \times 10^8) = 10 (8.70) \text{ dB} = 87 \text{ dB}.$$

Discussion

This 87 dB sound has an intensity five times as great as an 80 dB sound. So a factor of five in intensity corresponds to a difference of 7 dB in sound intensity level. This value is true for any intensities differing by a factor of five.

Example:**Change Intensity Levels of a Sound: What Happens to the Decibel Level?**

Show that if one sound is twice as intense as another, it has a sound level about 3 dB higher.

Strategy

You are given that the ratio of two intensities is 2 to 1, and are then asked to find the difference in their sound levels in decibels. You can solve this problem using the properties of logarithms.

Solution

(1) Identify knowns:

The ratio of the two intensities is 2 to 1, or:

Equation:

$$\frac{I_2}{I_1} = 2.00.$$

We wish to show that the difference in sound levels is about 3 dB. That is, we want to show:

Equation:

$$\beta_2 - \beta_1 = 3 \text{ dB}.$$

Note that:

Equation:

$$\log_{10}b - \log_{10}a = \log_{10}\left(\frac{b}{a}\right).$$

(2) Use the definition of β to get:

Equation:

$$\beta_2 - \beta_1 = 10 \log_{10} \left(\frac{I_2}{I_1} \right) = 10 \log_{10} 2.00 = 10 (0.301) \text{ dB}.$$

Thus,

Equation:

$$\beta_2 - \beta_1 = 3.01 \text{ dB}.$$

Discussion

This means that the two sound intensity levels differ by 3.01 dB, or about 3 dB, as advertised. Note that because only the ratio I_2/I_1 is given (and not the actual intensities), this result is true for any intensities that differ by a factor of two. For example, a 56.0 dB sound is twice as intense as a 53.0 dB sound, a 97.0 dB sound is half as intense as a 100 dB sound, and so on.

It should be noted at this point that there is another decibel scale in use, called the **sound pressure level**, based on the ratio of the pressure amplitude to a reference pressure. This scale is used particularly in applications where sound travels in water. It is beyond the scope of most introductory texts to treat this scale because it is not commonly used for sounds in air, but it is important to note that very different decibel levels may be encountered when sound pressure levels are quoted. For example, ocean noise pollution produced by ships may be as great as 200 dB expressed in the sound pressure level, where the more familiar sound intensity level we use here would be something under 140 dB for the same sound.

Note:

Take-Home Investigation: Feeling Sound

Find a CD player and a CD that has rock music. Place the player on a light table, insert the CD into the player, and start playing the CD. Place your hand gently on the table next to the speakers. Increase the volume and note the level when the table just begins to vibrate as the rock music plays. Increase the reading on the volume control until it doubles. What has happened to the vibrations?

Exercise:

Check Your Understanding

Problem:

Describe how amplitude is related to the loudness of a sound.

Solution:

Amplitude is directly proportional to the experience of loudness. As amplitude increases, loudness increases.

Exercise:

Check Your Understanding

Problem:

Identify common sounds at the levels of 10 dB, 50 dB, and 100 dB.

Solution:

10 dB: Running fingers through your hair.

50 dB: Inside a quiet home with no television or radio.

100 dB: Take-off of a jet plane.

Section Summary

- Intensity is the same for a sound wave as was defined for all waves; it is

Equation:

$$I = \frac{P}{A},$$

where P is the power crossing area A . The SI unit for I is watts per meter squared. The intensity of a sound wave is also related to the pressure amplitude Δp

Equation:

$$I = \frac{(\Delta p)^2}{2\rho v_w},$$

where ρ is the density of the medium in which the sound wave travels and v_w is the speed of sound in the medium.

- Sound intensity level in units of decibels (dB) is

Equation:

$$\beta \text{ (dB)} = 10 \log_{10} \left(\frac{I}{I_0} \right),$$

where $I_0 = 10^{-12} \text{ W/m}^2$ is the threshold intensity of hearing.

Conceptual Questions

Exercise:

Problem:

Six members of a synchronized swim team wear earplugs to protect themselves against water pressure at depths, but they can still hear the music and perform the combinations in the water perfectly. One day, they were asked to leave the pool so the dive team could practice a few dives, and they tried to practice on a mat, but seemed to have a lot more difficulty. Why might this be?

Exercise:**Problem:**

A community is concerned about a plan to bring train service to their downtown from the town's outskirts. The current sound intensity level, even though the rail yard is blocks away, is 70 dB downtown. The mayor assures the public that there will be a difference of only 30 dB in sound in the downtown area. Should the townspeople be concerned? Why?

Problems & Exercises**Exercise:****Problem:**

What is the intensity in watts per meter squared of 85.0-dB sound?

Solution:**Equation:**

$$3.16 \times 10^{-4} \text{ W/m}^2$$

Exercise:

Problem:

The warning tag on a lawn mower states that it produces noise at a level of 91.0 dB. What is this in watts per meter squared?

Exercise:**Problem:**

A sound wave traveling in 20°C air has a pressure amplitude of 0.5 Pa. What is the intensity of the wave?

Solution:**Equation:**

$$3.04 \times 10^{-4} \text{ W/m}^2$$

Exercise:**Problem:**

What intensity level does the sound in the preceding problem correspond to?

Exercise:**Problem:**

What sound intensity level in dB is produced by earphones that create an intensity of $4.00 \times 10^{-2} \text{ W/m}^2$?

Solution:

106 dB

Exercise:**Problem:**

Show that an intensity of 10^{-12} W/m^2 is the same as 10^{-16} W/cm^2 .

Exercise:**Problem:**

(a) What is the decibel level of a sound that is twice as intense as a 90.0-dB sound? (b) What is the decibel level of a sound that is one-fifth as intense as a 90.0-dB sound?

Solution:

(a) 93 dB

(b) 83 dB

Exercise:**Problem:**

(a) What is the intensity of a sound that has a level 7.00 dB lower than a $4.00 \times 10^{-9} \text{ W/m}^2$ sound? (b) What is the intensity of a sound that is 3.00 dB higher than a $4.00 \times 10^{-9} \text{ W/m}^2$ sound?

Exercise:**Problem:**

(a) How much more intense is a sound that has a level 17.0 dB higher than another? (b) If one sound has a level 23.0 dB less than another, what is the ratio of their intensities?

Solution:

(a) 50.1

(b) 5.01×10^{-3} or $\frac{1}{200}$

Exercise:

Problem:

People with good hearing can perceive sounds as low in level as -8.00 dB at a frequency of 3000 Hz. What is the intensity of this sound in watts per meter squared?

Exercise:**Problem:**

If a large housefly 3.0 m away from you makes a noise of 40.0 dB, what is the noise level of 1000 flies at that distance, assuming interference has a negligible effect?

Solution:

70.0 dB

Exercise:**Problem:**

Ten cars in a circle at a boom box competition produce a 120 -dB sound intensity level at the center of the circle. What is the average sound intensity level produced there by each stereo, assuming interference effects can be neglected?

Exercise:**Problem:**

The amplitude of a sound wave is measured in terms of its maximum gauge pressure. By what factor does the amplitude of a sound wave increase if the sound intensity level goes up by 40.0 dB?

Solution:

100

Exercise:

Problem:

If a sound intensity level of 0 dB at 1000 Hz corresponds to a maximum gauge pressure (sound amplitude) of 10^{-9} atm, what is the maximum gauge pressure in a 60-dB sound? What is the maximum gauge pressure in a 120-dB sound?

Exercise:**Problem:**

An 8-hour exposure to a sound intensity level of 90.0 dB may cause hearing damage. What energy in joules falls on a 0.800-cm-diameter eardrum so exposed?

Solution:**Equation:**

$$1.45 \times 10^{-3} \text{ J}$$

Exercise:**Problem:**

(a) Ear trumpets were never very common, but they did aid people with hearing losses by gathering sound over a large area and concentrating it on the smaller area of the eardrum. What decibel increase does an ear trumpet produce if its sound gathering area is 900 cm^2 and the area of the eardrum is 0.500 cm^2 , but the trumpet only has an efficiency of 5.00% in transmitting the sound to the eardrum? (b) Comment on the usefulness of the decibel increase found in part (a).

Exercise:

Problem:

Sound is more effectively transmitted into a stethoscope by direct contact than through the air, and it is further intensified by being concentrated on the smaller area of the eardrum. It is reasonable to assume that sound is transmitted into a stethoscope 100 times as effectively compared with transmission through the air. What, then, is the gain in decibels produced by a stethoscope that has a sound gathering area of 15.0 cm^2 , and concentrates the sound onto two eardrums with a total area of 0.900 cm^2 with an efficiency of 40.0%?

Solution:

28.2 dB

Exercise:**Problem:**

Loudspeakers can produce intense sounds with surprisingly small energy input in spite of their low efficiencies. Calculate the power input needed to produce a 90.0-dB sound intensity level for a 12.0-cm-diameter speaker that has an efficiency of 1.00%. (This value is the sound intensity level right at the speaker.)

Glossary

intensity

the power per unit area carried by a wave

sound intensity level

a unitless quantity telling you the level of the sound relative to a fixed standard

sound pressure level

the ratio of the pressure amplitude to a reference pressure

Doppler Effect and Sonic Booms

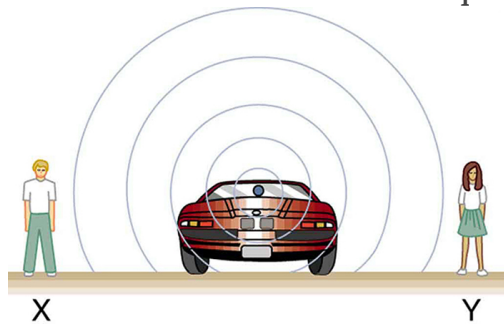
- Define Doppler effect, Doppler shift, and sonic boom.
- Calculate the frequency of a sound heard by someone observing Doppler shift.
- Describe the sounds produced by objects moving faster than the speed of sound.

The characteristic sound of a motorcycle buzzing by is an example of the **Doppler effect**. The high-pitch scream shifts dramatically to a lower-pitch roar as the motorcycle passes by a stationary observer. The closer the motorcycle brushes by, the more abrupt the shift. The faster the motorcycle moves, the greater the shift. We also hear this characteristic shift in frequency for passing race cars, airplanes, and trains. It is so familiar that it is used to imply motion and children often mimic it in play.

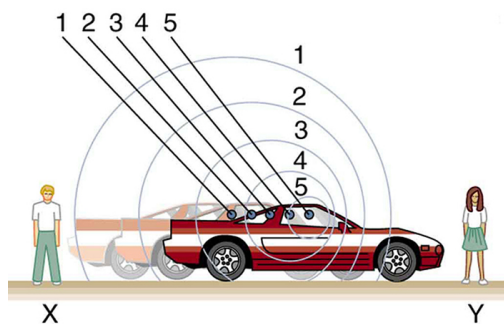
The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer. Although less familiar, this effect is easily noticed for a stationary source and moving observer. For example, if you ride a train past a stationary warning bell, you will hear the bell's frequency shift from high to low as you pass by. The actual change in frequency due to relative motion of source and observer is called a **Doppler shift**. The Doppler effect and Doppler shift are named for the Austrian physicist and mathematician Christian Johann Doppler (1803–1853), who did experiments with both moving sources and moving observers. Doppler, for example, had musicians play on a moving open train car and also play standing next to the train tracks as a train passed by. Their music was observed both on and off the train, and changes in frequency were measured.

What causes the Doppler shift? [\[link\]](#), [\[link\]](#), and [\[link\]](#) compare sound waves emitted by stationary and moving sources in a stationary air mass. Each disturbance spreads out spherically from the point where the sound was emitted. If the source is stationary, then all of the spheres representing the air compressions in the sound wave centered on the same point, and the stationary observers on either side see the same wavelength and frequency as emitted by the source, as in [\[link\]](#). If the source is moving, as in [\[link\]](#), then the situation is different. Each compression of the air moves out in a

sphere from the point where it was emitted, but the point of emission moves. This moving emission point causes the air compressions to be closer together on one side and farther apart on the other. Thus, the wavelength is shorter in the direction the source is moving (on the right in [\[link\]](#)), and longer in the opposite direction (on the left in [\[link\]](#)). Finally, if the observers move, as in [\[link\]](#), the frequency at which they receive the compressions changes. The observer moving toward the source receives them at a higher frequency, and the person moving away from the source receives them at a lower frequency.

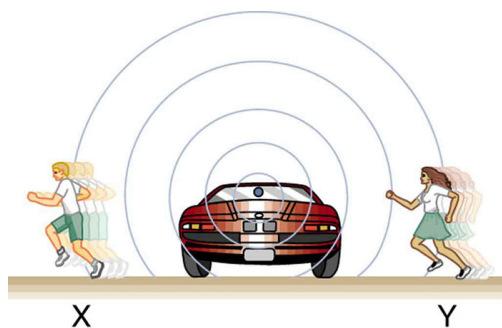


Sounds emitted by a source spread out in spherical waves. Because the source, observers, and air are stationary, the wavelength and frequency are the same in all directions and to all observers.



Sounds emitted by a

source moving to the right spread out from the points at which they were emitted. The wavelength is reduced and, consequently, the frequency is increased in the direction of motion, so that the observer on the right hears a higher-pitch sound. The opposite is true for the observer on the left, where the wavelength is increased and the frequency is reduced.



The same effect is produced when the observers move relative to the source. Motion toward the source increases frequency as the observer on the right passes through more wave crests than she would if stationary. Motion away from the

source decreases frequency as the observer on the left passes through fewer wave crests than he would if stationary.

We know that wavelength and frequency are related by $v_w = f\lambda$, where v_w is the fixed speed of sound. The sound moves in a medium and has the same speed v_w in that medium whether the source is moving or not. Thus f multiplied by λ is a constant. Because the observer on the right in [\[link\]](#) receives a shorter wavelength, the frequency she receives must be higher. Similarly, the observer on the left receives a longer wavelength, and hence he hears a lower frequency. The same thing happens in [\[link\]](#). A higher frequency is received by the observer moving toward the source, and a lower frequency is received by an observer moving away from the source. In general, then, relative motion of source and observer toward one another increases the received frequency. Relative motion apart decreases frequency. The greater the relative speed is, the greater the effect.

Note:**The Doppler Effect**

The Doppler effect occurs not only for sound but for any wave when there is relative motion between the observer and the source. There are Doppler shifts in the frequency of sound, light, and water waves, for example.

Doppler shifts can be used to determine velocity, such as when ultrasound is reflected from blood in a medical diagnostic. The recession of galaxies is determined by the shift in the frequencies of light received from them and has implied much about the origins of the universe. Modern physics has been profoundly affected by observations of Doppler shifts.

For a stationary observer and a moving source, the frequency f_{obs} received by the observer can be shown to be

Equation:

$$f_{\text{obs}} = f_s \left(\frac{v_w}{v_w \pm v_s} \right),$$

where f_s is the frequency of the source, v_s is the speed of the source along a line joining the source and observer, and v_w is the speed of sound. The minus sign is used for motion toward the observer and the plus sign for motion away from the observer, producing the appropriate shifts up and down in frequency. Note that the greater the speed of the source, the greater the effect. Similarly, for a stationary source and moving observer, the frequency received by the observer f_{obs} is given by

Equation:

$$f_{\text{obs}} = f_s \left(\frac{v_w \pm v_{\text{obs}}}{v_w} \right),$$

where v_{obs} is the speed of the observer along a line joining the source and observer. Here the plus sign is for motion toward the source, and the minus is for motion away from the source.

Example:**Calculate Doppler Shift: A Train Horn**

Suppose a train that has a 150-Hz horn is moving at 35.0 m/s in still air on a day when the speed of sound is 340 m/s.

(a) What frequencies are observed by a stationary person at the side of the tracks as the train approaches and after it passes?

(b) What frequency is observed by the train's engineer traveling on the train?

Strategy

To find the observed frequency in (a), $f_{\text{obs}} = f_s \left(\frac{v_w}{v_w \pm v_s} \right)$, must be used because the source is moving. The minus sign is used for the approaching

train, and the plus sign for the receding train. In (b), there are two Doppler shifts—one for a moving source and the other for a moving observer.

Solution for (a)

(1) Enter known values into $f_{\text{obs}} = f_s \left(\frac{v_w}{v_w - v_s} \right)$.

Equation:

$$f_{\text{obs}} = f_s \left(\frac{v_w}{v_w - v_s} \right) = (150 \text{ Hz}) \left(\frac{340 \text{ m/s}}{340 \text{ m/s} - 35.0 \text{ m/s}} \right)$$

(2) Calculate the frequency observed by a stationary person as the train approaches.

Equation:

$$f_{\text{obs}} = (150 \text{ Hz})(1.11) = 167 \text{ Hz}$$

(3) Use the same equation with the plus sign to find the frequency heard by a stationary person as the train recedes.

Equation:

$$f_{\text{obs}} = f_s \left(\frac{v_w}{v_w + v_s} \right) = (150 \text{ Hz}) \left(\frac{340 \text{ m/s}}{340 \text{ m/s} + 35.0 \text{ m/s}} \right)$$

(4) Calculate the second frequency.

Equation:

$$f_{\text{obs}} = (150 \text{ Hz})(0.907) = 136 \text{ Hz}$$

Discussion on (a)

The numbers calculated are valid when the train is far enough away that the motion is nearly along the line joining train and observer. In both cases, the shift is significant and easily noticed. Note that the shift is 17.0 Hz for motion toward and 14.0 Hz for motion away. The shifts are not symmetric.

Solution for (b)

(1) Identify knowns:

- It seems reasonable that the engineer would receive the same frequency as emitted by the horn, because the relative velocity

between them is zero.

- Relative to the medium (air), the speeds are $v_s = v_{\text{obs}} = 35.0 \text{ m/s}$.
- The first Doppler shift is for the moving observer; the second is for the moving source.

(2) Use the following equation:

Equation:

$$f_{\text{obs}} = \left[f_s \left(\frac{v_w \pm v_{\text{obs}}}{v_w} \right) \right] \left(\frac{v_w}{v_w \pm v_s} \right).$$

The quantity in the square brackets is the Doppler-shifted frequency due to a moving observer. The factor on the right is the effect of the moving source.

(3) Because the train engineer is moving in the direction toward the horn, we must use the plus sign for v_{obs} ; however, because the horn is also moving in the direction away from the engineer, we also use the plus sign for v_s . But the train is carrying both the engineer and the horn at the same velocity, so $v_s = v_{\text{obs}}$. As a result, everything but f_s cancels, yielding

Equation:

$$f_{\text{obs}} = f_s.$$

Discussion for (b)

We may expect that there is no change in frequency when source and observer move together because it fits your experience. For example, there is no Doppler shift in the frequency of conversations between driver and passenger on a motorcycle. People talking when a wind moves the air between them also observe no Doppler shift in their conversation. The crucial point is that source and observer are not moving relative to each other.

Sonic Booms to Bow Wakes

What happens to the sound produced by a moving source, such as a jet airplane, that approaches or even exceeds the speed of sound? The answer

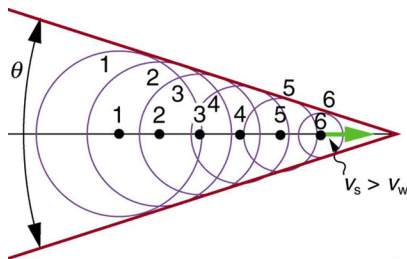
to this question applies not only to sound but to all other waves as well.

Suppose a jet airplane is coming nearly straight at you, emitting a sound of frequency f_s . The greater the plane's speed v_s , the greater the Doppler shift and the greater the value observed for f_{obs} . Now, as v_s approaches the speed of sound, f_{obs} approaches infinity, because the denominator in

$f_{\text{obs}} = f_s \left(\frac{v_w}{v_w \pm v_s} \right)$ approaches zero. At the speed of sound, this result

means that in front of the source, each successive wave is superimposed on the previous one because the source moves forward at the speed of sound.

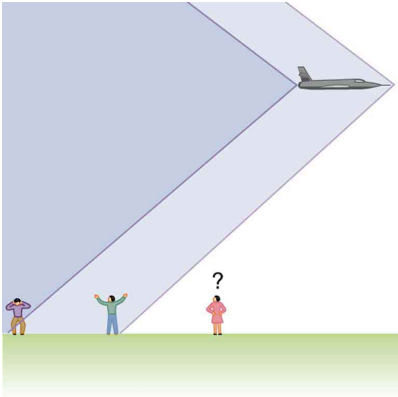
The observer gets them all at the same instant, and so the frequency is infinite. (Before airplanes exceeded the speed of sound, some people argued it would be impossible because such constructive superposition would produce pressures great enough to destroy the airplane.) If the source exceeds the speed of sound, no sound is received by the observer until the source has passed, so that the sounds from the approaching source are mixed with those from it when receding. This mixing appears messy, but something interesting happens—a sonic boom is created. (See [\[link\]](#).)



Sound waves from
a source that moves
faster than the
speed of sound
spread spherically
from the point
where they are
emitted, but the
source moves
ahead of each.

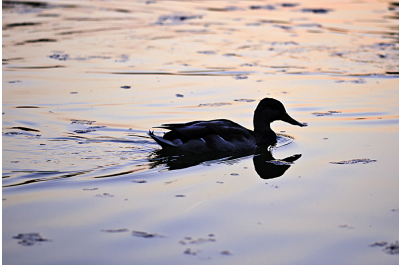
Constructive
interference along
the lines shown
(actually a cone in
three dimensions)
creates a shock
wave called a sonic
boom. The faster
the speed of the
source, the smaller
the angle θ .

There is constructive interference along the lines shown (a cone in three dimensions) from similar sound waves arriving there simultaneously. This superposition forms a disturbance called a **sonic boom**, a constructive interference of sound created by an object moving faster than sound. Inside the cone, the interference is mostly destructive, and so the sound intensity there is much less than on the shock wave. An aircraft creates two sonic booms, one from its nose and one from its tail. (See [\[link\]](#).) During television coverage of space shuttle landings, two distinct booms could often be heard. These were separated by exactly the time it would take the shuttle to pass by a point. Observers on the ground often do not see the aircraft creating the sonic boom, because it has passed by before the shock wave reaches them, as seen in [\[link\]](#). If the aircraft flies close by at low altitude, pressures in the sonic boom can be destructive and break windows as well as rattle nerves. Because of how destructive sonic booms can be, supersonic flights are banned over populated areas of the United States.

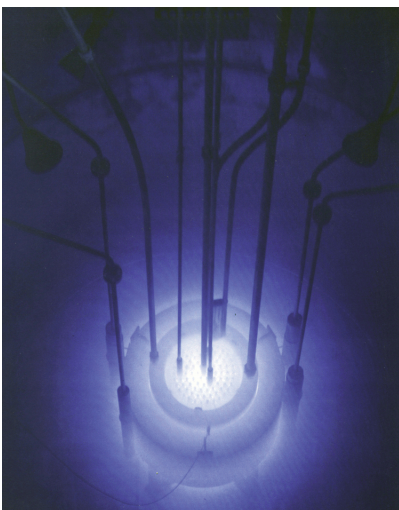


Two sonic booms,
created by the nose
and tail of an
aircraft, are
observed on the
ground after the
plane has passed
by.

Sonic booms are one example of a broader phenomenon called bow wakes. A **bow wake**, such as the one in [\[link\]](#), is created when the wave source moves faster than the wave propagation speed. Water waves spread out in circles from the point where created, and the bow wake is the familiar V-shaped wake trailing the source. A more exotic bow wake is created when a subatomic particle travels through a medium faster than the speed of light travels in that medium. (In a vacuum, the maximum speed of light will be $c = 3.00 \times 10^8$ m/s; in the medium of water, the speed of light is closer to $0.75c$. If the particle creates light in its passage, that light spreads on a cone with an angle indicative of the speed of the particle, as illustrated in [\[link\]](#). Such a bow wake is called Cerenkov radiation and is commonly observed in particle physics.



Bow wake created
by a duck.
Constructive
interference
produces the rather
structured wake,
while there is
relatively little
wave action inside
the wake, where
interference is
mostly destructive.
(credit: Horia
Varlan, Flickr)



The blue glow in
this research
reactor pool is
Cerenkov radiation
caused by
subatomic particles
traveling faster than
the speed of light in
water. (credit: U.S.
Nuclear Regulatory
Commission)

Doppler shifts and sonic booms are interesting sound phenomena that occur in all types of waves. They can be of considerable use. For example, the Doppler shift in ultrasound can be used to measure blood velocity, while police use the Doppler shift in radar (a microwave) to measure car velocities. In meteorology, the Doppler shift is used to track the motion of storm clouds; such “Doppler Radar” can give velocity and direction and rain or snow potential of imposing weather fronts. In astronomy, we can examine the light emitted from distant galaxies and determine their speed relative to ours. As galaxies move away from us, their light is shifted to a lower frequency, and so to a longer wavelength—the so-called red shift. Such information from galaxies far, far away has allowed us to estimate the age of the universe (from the Big Bang) as about 14 billion years.

Exercise:

Check Your Understanding

Problem:

Why did scientist Christian Doppler observe musicians both on a moving train and also from a stationary point not on the train?

Solution:

Doppler needed to compare the perception of sound when the observer is stationary and the sound source moves, as well as when the sound

source and the observer are both in motion.

Exercise:

Check Your Understanding

Problem:

Describe a situation in your life when you might rely on the Doppler shift to help you either while driving a car or walking near traffic.

Solution:

If I am driving and I hear Doppler shift in an ambulance siren, I would be able to tell when it was getting closer and also if it has passed by. This would help me to know whether I needed to pull over and let the ambulance through.

Section Summary

- The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer.
- The actual change in frequency is called the Doppler shift.
- A sonic boom is constructive interference of sound created by an object moving faster than sound.
- A sonic boom is a type of bow wake created when any wave source moves faster than the wave propagation speed.
- For a stationary observer and a moving source, the observed frequency f_{obs} is:

Equation:

$$f_{\text{obs}} = f_s \left(\frac{v_w}{v_w \pm v_s} \right),$$

where f_s is the frequency of the source, v_s is the speed of the source, and v_w is the speed of sound. The minus sign is used for motion toward the observer and the plus sign for motion away.

- For a stationary source and moving observer, the observed frequency is:

Equation:

$$f_{\text{obs}} = f_s \left(\frac{v_w \pm v_{\text{obs}}}{v_w} \right),$$

where v_{obs} is the speed of the observer.

Conceptual Questions

Exercise:

Problem: Is the Doppler shift real or just a sensory illusion?

Exercise:

Problem:

Due to efficiency considerations related to its bow wake, the supersonic transport aircraft must maintain a cruising speed that is a constant ratio to the speed of sound (a constant Mach number). If the aircraft flies from warm air into colder air, should it increase or decrease its speed? Explain your answer.

Exercise:

Problem:

When you hear a sonic boom, you often cannot see the plane that made it. Why is that?

Problems & Exercises

Exercise:

Problem:

(a) What frequency is received by a person watching an oncoming ambulance moving at 110 km/h and emitting a steady 800-Hz sound from its siren? The speed of sound on this day is 345 m/s. (b) What frequency does she receive after the ambulance has passed?

Solution:

(a) 878 Hz

(b) 735 Hz

Exercise:**Problem:**

(a) At an air show a jet flies directly toward the stands at a speed of 1200 km/h, emitting a frequency of 3500 Hz, on a day when the speed of sound is 342 m/s. What frequency is received by the observers? (b) What frequency do they receive as the plane flies directly away from them?

Exercise:**Problem:**

What frequency is received by a mouse just before being dispatched by a hawk flying at it at 25.0 m/s and emitting a screech of frequency 3500 Hz? Take the speed of sound to be 331 m/s.

Solution:**Equation:**

$$3.79 \times 10^3 \text{ Hz}$$

Exercise:

Problem:

A spectator at a parade receives an 888-Hz tone from an oncoming trumpeter who is playing an 880-Hz note. At what speed is the musician approaching if the speed of sound is 338 m/s?

Exercise:**Problem:**

A commuter train blows its 200-Hz horn as it approaches a crossing. The speed of sound is 335 m/s. (a) An observer waiting at the crossing receives a frequency of 208 Hz. What is the speed of the train? (b) What frequency does the observer receive as the train moves away?

Solution:

(a) 12.9 m/s

(b) 193 Hz

Exercise:**Problem:**

Can you perceive the shift in frequency produced when you pull a tuning fork toward you at 10.0 m/s on a day when the speed of sound is 344 m/s? To answer this question, calculate the factor by which the frequency shifts and see if it is greater than 0.300%.

Exercise:**Problem:**

Two eagles fly directly toward one another, the first at 15.0 m/s and the second at 20.0 m/s. Both screech, the first one emitting a frequency of 3200 Hz and the second one emitting a frequency of 3800 Hz. What frequencies do they receive if the speed of sound is 330 m/s?

Solution:

First eagle hears $4.23 \times 10^3 \text{ Hz}$

Second eagle hears $3.56 \times 10^3 \text{ Hz}$

Exercise:

Problem:

What is the minimum speed at which a source must travel toward you for you to be able to hear that its frequency is Doppler shifted? That is, what speed produces a shift of 0.300% on a day when the speed of sound is 331 m/s?

Glossary

Doppler effect

an alteration in the observed frequency of a sound due to motion of either the source or the observer

Doppler shift

the actual change in frequency due to relative motion of source and observer

sonic boom

a constructive interference of sound created by an object moving faster than sound

bow wake

V-shaped disturbance created when the wave source moves faster than the wave propagation speed

Sound Interference and Resonance: Standing Waves in Air Columns

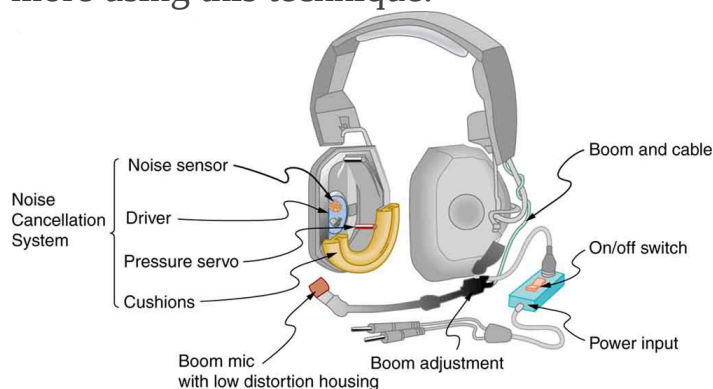
- Define antinode, node, fundamental, overtones, and harmonics.
- Identify instances of sound interference in everyday situations.
- Describe how sound interference occurring inside open and closed tubes changes the characteristics of the sound, and how this applies to sounds produced by musical instruments.
- Calculate the length of a tube using sound wave measurements.



Some types
of
headphones
use the
phenomena
of
constructiv
e and
destructive
interference
to cancel
out outside
noises.
(credit:
JVC
America,
Flickr)

Interference is the hallmark of waves, all of which exhibit constructive and destructive interference exactly analogous to that seen for water waves. In fact, one way to prove something “is a wave” is to observe interference effects. So, sound being a wave, we expect it to exhibit interference; we have already mentioned a few such effects, such as the beats from two similar notes played simultaneously.

[\[link\]](#) shows a clever use of sound interference to cancel noise. Larger-scale applications of active noise reduction by destructive interference are contemplated for entire passenger compartments in commercial aircraft. To obtain destructive interference, a fast electronic analysis is performed, and a second sound is introduced with its maxima and minima exactly reversed from the incoming noise. Sound waves in fluids are pressure waves and consistent with Pascal’s principle; pressures from two different sources add and subtract like simple numbers; that is, positive and negative gauge pressures add to a much smaller pressure, producing a lower-intensity sound. Although completely destructive interference is possible only under the simplest conditions, it is possible to reduce noise levels by 30 dB or more using this technique.



Headphones designed to cancel noise with destructive interference create a sound wave exactly opposite to the incoming sound. These headphones can be more effective than the simple passive attenuation used in most ear protection. Such headphones were

used on the record-setting, around the world nonstop flight of the Voyager aircraft to protect the pilots' hearing from engine noise.

Where else can we observe sound interference? All sound resonances, such as in musical instruments, are due to constructive and destructive interference. Only the resonant frequencies interfere constructively to form standing waves, while others interfere destructively and are absent. From the toot made by blowing over a bottle, to the characteristic flavor of a violin's sounding box, to the recognizability of a great singer's voice, resonance and standing waves play a vital role.

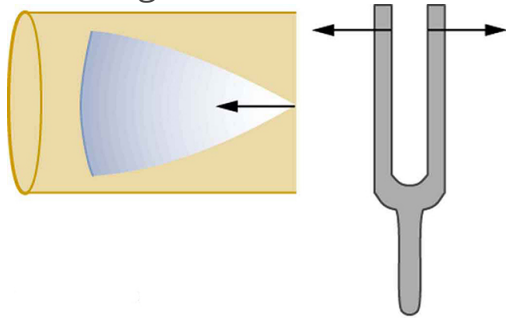
Note:

Interference

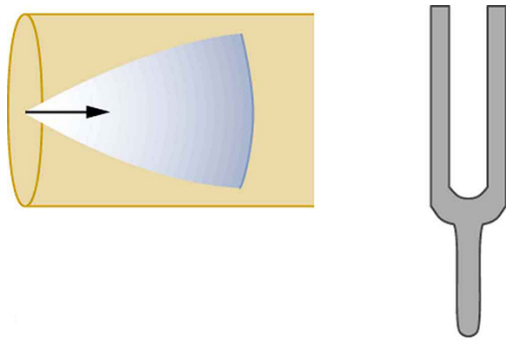
Interference is such a fundamental aspect of waves that observing interference is proof that something is a wave. The wave nature of light was established by experiments showing interference. Similarly, when electrons scattered from crystals exhibited interference, their wave nature was confirmed to be exactly as predicted by symmetry with certain wave characteristics of light.

Suppose we hold a tuning fork near the end of a tube that is closed at the other end, as shown in [\[link\]](#), [\[link\]](#), [\[link\]](#), and [\[link\]](#). If the tuning fork has just the right frequency, the air column in the tube resonates loudly, but at most frequencies it vibrates very little. This observation just means that the air column has only certain natural frequencies. The figures show how a resonance at the lowest of these natural frequencies is formed. A disturbance travels down the tube at the speed of sound and bounces off the closed end. If the tube is just the right length, the reflected sound arrives back at the tuning fork exactly half a cycle later, and it interferes

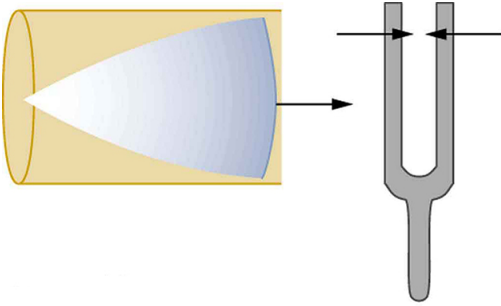
constructively with the continuing sound produced by the tuning fork. The incoming and reflected sounds form a standing wave in the tube as shown.



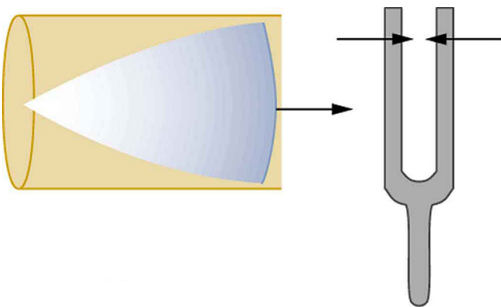
Resonance of air in a tube closed at one end, caused by a tuning fork. A disturbance moves down the tube.



Resonance of air in a tube closed at one end, caused by a tuning fork. The disturbance reflects from the closed end of the tube.



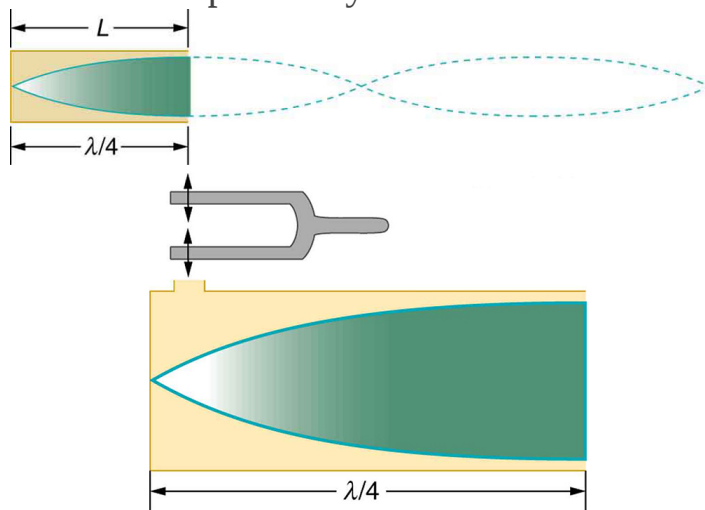
Resonance of air in a tube closed at one end, caused by a tuning fork. If the length of the tube L is just right, the disturbance gets back to the tuning fork half a cycle later and interferes constructively with the continuing sound from the tuning fork. This interference forms a standing wave, and the air column resonates.



Resonance of air in a tube closed at one end, caused by a tuning fork. A graph of air displacement along the length of the tube shows none at the closed

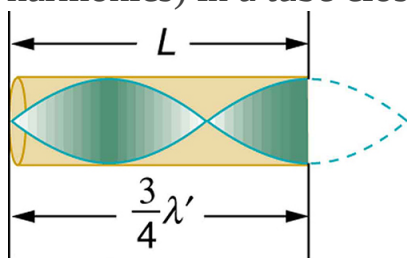
end, where the motion is constrained, and a maximum at the open end. This standing wave has one-fourth of its wavelength in the tube, so that $\lambda = 4L$.

The standing wave formed in the tube has its maximum air displacement (an **antinode**) at the open end, where motion is unconstrained, and no displacement (a **node**) at the closed end, where air movement is halted. The distance from a node to an antinode is one-fourth of a wavelength, and this equals the length of the tube; thus, $\lambda = 4L$. This same resonance can be produced by a vibration introduced at or near the closed end of the tube, as shown in [\[link\]](#). It is best to consider this a natural vibration of the air column independently of how it is induced.

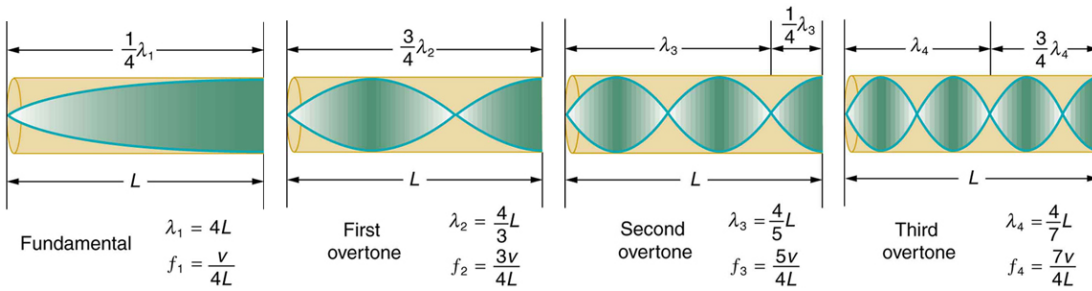


The same standing wave is created in the tube by a vibration introduced near its closed end.

Given that maximum air displacements are possible at the open end and none at the closed end, there are other, shorter wavelengths that can resonate in the tube, such as the one shown in [\[link\]](#). Here the standing wave has three-fourths of its wavelength in the tube, or $L = (3/4)\lambda'$, so that $\lambda' = 4L/3$. Continuing this process reveals a whole series of shorter-wavelength and higher-frequency sounds that resonate in the tube. We use specific terms for the resonances in any system. The lowest resonant frequency is called the **fundamental**, while all higher resonant frequencies are called **overtones**. All resonant frequencies are integral multiples of the fundamental, and they are collectively called **harmonics**. The fundamental is the first harmonic, the first overtone is the second harmonic, and so on. [\[link\]](#) shows the fundamental and the first three overtones (the first four harmonics) in a tube closed at one end.

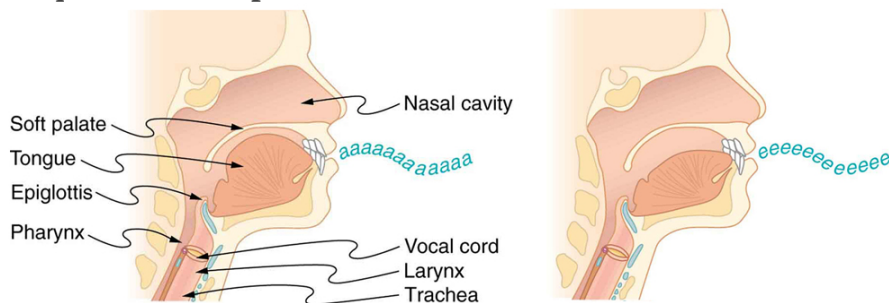


Another resonance for a tube closed at one end. This has maximum air displacements at the open end, and none at the closed end. The wavelength is shorter, with three-fourths λ' equaling the length of the tube, so that $\lambda' = 4L/3$. This higher-frequency vibration is the first overtone.



The fundamental and three lowest overtones for a tube closed at one end. All have maximum air displacements at the open end and none at the closed end.

The fundamental and overtones can be present simultaneously in a variety of combinations. For example, middle C on a trumpet has a sound distinctively different from middle C on a clarinet, both instruments being modified versions of a tube closed at one end. The fundamental frequency is the same (and usually the most intense), but the overtones and their mix of intensities are different and subject to shading by the musician. This mix is what gives various musical instruments (and human voices) their distinctive characteristics, whether they have air columns, strings, sounding boxes, or drumheads. In fact, much of our speech is determined by shaping the cavity formed by the throat and mouth and positioning the tongue to adjust the fundamental and combination of overtones. Simple resonant cavities can be made to resonate with the sound of the vowels, for example. (See [\[link\]](#).) In boys, at puberty, the larynx grows and the shape of the resonant cavity changes giving rise to the difference in predominant frequencies in speech between men and women.



The throat and mouth form an air column closed at one end that resonates in response to vibrations in the voice box. The spectrum of overtones and their intensities vary with mouth shaping and tongue position to form different sounds. The voice box can be replaced with a mechanical vibrator, and understandable speech is still possible. Variations in basic shapes make different voices recognizable.

Now let us look for a pattern in the resonant frequencies for a simple tube that is closed at one end. The fundamental has $\lambda = 4L$, and frequency is related to wavelength and the speed of sound as given by:

Equation:

$$v_w = f\lambda.$$

Solving for f in this equation gives

Equation:

$$f = \frac{v_w}{\lambda} = \frac{v_w}{4L},$$

where v_w is the speed of sound in air. Similarly, the first overtone has $\lambda' = 4L/3$ (see [\[link\]](#)), so that

Equation:

$$f' = 3 \frac{v_w}{4L} = 3f.$$

Because $f' = 3f$, we call the first overtone the third harmonic. Continuing this process, we see a pattern that can be generalized in a single expression. The resonant frequencies of a tube closed at one end are

Equation:

$$f_n = n \frac{v_w}{4L}, n = 1, 3, 5,$$

where f_1 is the fundamental, f_3 is the first overtone, and so on. It is interesting that the resonant frequencies depend on the speed of sound and, hence, on temperature. This dependence poses a noticeable problem for organs in old unheated cathedrals, and it is also the reason why musicians commonly bring their wind instruments to room temperature before playing them.

Example:

Find the Length of a Tube with a 128 Hz Fundamental

(a) What length should a tube closed at one end have on a day when the air temperature, is 22.0°C , if its fundamental frequency is to be 128 Hz (C below middle C)?

(b) What is the frequency of its fourth overtone?

Strategy

The length L can be found from the relationship in $f_n = n \frac{v_w}{4L}$, but we will first need to find the speed of sound v_w .

Solution for (a)

(1) Identify knowns:

- the fundamental frequency is 128 Hz
- the air temperature is 22.0°C

(2) Use $f_n = n \frac{v_w}{4L}$ to find the fundamental frequency ($n = 1$).

Equation:

$$f_1 = \frac{v_w}{4L}$$

(3) Solve this equation for length.

Equation:

$$L = \frac{v_w}{4f_1}$$

(4) Find the speed of sound using $v_w = (331 \text{ m/s}) \sqrt{\frac{T}{273 \text{ K}}}$.

Equation:

$$v_w = (331 \text{ m/s}) \sqrt{\frac{295 \text{ K}}{273 \text{ K}}} = 344 \text{ m/s}$$

(5) Enter the values of the speed of sound and frequency into the expression for L .

Equation:

$$L = \frac{v_w}{4f_1} = \frac{344 \text{ m/s}}{4(128 \text{ Hz})} = 0.672 \text{ m}$$

Discussion on (a)

Many wind instruments are modified tubes that have finger holes, valves, and other devices for changing the length of the resonating air column and hence, the frequency of the note played. Horns producing very low frequencies, such as tubas, require tubes so long that they are coiled into loops.

Solution for (b)

(1) Identify knowns:

- the first overtone has $n = 3$
- the second overtone has $n = 5$
- the third overtone has $n = 7$
- the fourth overtone has $n = 9$

(2) Enter the value for the fourth overtone into $f_n = n \frac{v_w}{4L}$.

Equation:

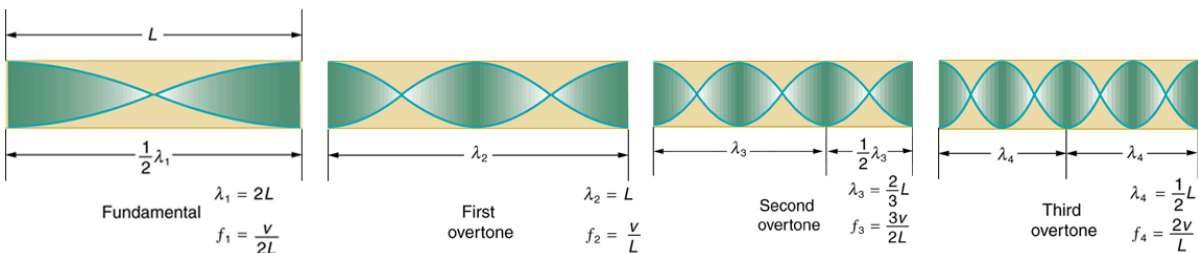
$$f_9 = 9 \frac{v_w}{4L} = 9f_1 = 1.15 \text{ kHz}$$

Discussion on (b)

Whether this overtone occurs in a simple tube or a musical instrument depends on how it is stimulated to vibrate and the details of its shape. The

trombone, for example, does not produce its fundamental frequency and only makes overtones.

Another type of tube is one that is *open* at both ends. Examples are some organ pipes, flutes, and oboes. The resonances of tubes open at both ends can be analyzed in a very similar fashion to those for tubes closed at one end. The air columns in tubes open at both ends have maximum air displacements at both ends, as illustrated in [\[link\]](#). Standing waves form as shown.



The resonant frequencies of a tube open at both ends are shown, including the fundamental and the first three overtones. In all cases the maximum air displacements occur at both ends of the tube, giving it different natural frequencies than a tube closed at one end.

Based on the fact that a tube open at both ends has maximum air displacements at both ends, and using [\[link\]](#) as a guide, we can see that the resonant frequencies of a tube open at both ends are:

Equation:

$$f_n = n \frac{v_w}{2L}, \quad n = 1, 2, 3, \dots,$$

where f_1 is the fundamental, f_2 is the first overtone, f_3 is the second overtone, and so on. Note that a tube open at both ends has a fundamental frequency twice what it would have if closed at one end. It also has a different spectrum of overtones than a tube closed at one end. So if you had

two tubes with the same fundamental frequency but one was open at both ends and the other was closed at one end, they would sound different when played because they have different overtones. Middle C, for example, would sound richer played on an open tube, because it has even multiples of the fundamental as well as odd. A closed tube has only odd multiples.

Note:

Real-World Applications: Resonance in Everyday Systems

Resonance occurs in many different systems, including strings, air columns, and atoms. Resonance is the driven or forced oscillation of a system at its natural frequency. At resonance, energy is transferred rapidly to the oscillating system, and the amplitude of its oscillations grows until the system can no longer be described by Hooke's law. An example of this is the distorted sound intentionally produced in certain types of rock music.

Wind instruments use resonance in air columns to amplify tones made by lips or vibrating reeds. Other instruments also use air resonance in clever ways to amplify sound. [\[link\]](#) shows a violin and a guitar, both of which have sounding boxes but with different shapes, resulting in different overtone structures. The vibrating string creates a sound that resonates in the sounding box, greatly amplifying the sound and creating overtones that give the instrument its characteristic flavor. The more complex the shape of the sounding box, the greater its ability to resonate over a wide range of frequencies. The marimba, like the one shown in [\[link\]](#) uses pots or gourds below the wooden slats to amplify their tones. The resonance of the pot can be adjusted by adding water.



String instruments such as violins and guitars use resonance in their sounding boxes to amplify and enrich the sound created by their vibrating strings. The bridge and supports couple the string vibrations to the sounding boxes and air within.
(credits: guitar, Feliciano Guimares, Fotopedia; violin, Steve Snodgrass, Flickr)



Resonance has been used in musical instruments since prehistoric times. This marimba uses gourds as resonance chambers to amplify its sound.
(credit: APC Events, Flickr)

We have emphasized sound applications in our discussions of resonance and standing waves, but these ideas apply to any system that has wave characteristics. Vibrating strings, for example, are actually resonating and have fundamentals and overtones similar to those for air columns. More subtle are the resonances in atoms due to the wave character of their electrons. Their orbitals can be viewed as standing waves, which have a fundamental (ground state) and overtones (excited states). It is fascinating that wave characteristics apply to such a wide range of physical systems.

Exercise:

Check Your Understanding

Problem:

Describe how noise-canceling headphones differ from standard headphones used to block outside sounds.

Solution:

Regular headphones only block sound waves with a physical barrier. Noise-canceling headphones use destructive interference to reduce the loudness of outside sounds.

Exercise:

Check Your Understanding

Problem:

How is it possible to use a standing wave's node and antinode to determine the length of a closed-end tube?

Solution:

When the tube resonates at its natural frequency, the wave's node is located at the closed end of the tube, and the antinode is located at the open end. The length of the tube is equal to one-fourth of the wavelength of this wave. Thus, if we know the wavelength of the wave, we can determine the length of the tube.

Note:

PhET Explorations: Sound

This simulation lets you see sound waves. Adjust the frequency or volume and you can see and hear how the wave changes. Move the listener around and hear what she hears.

<https://archive.cnx.org/specials/c4d3b96e-41f3-11e5-ab7b-47e22dffc18e/sound/#sim-single-source>

Section Summary

- Sound interference and resonance have the same properties as defined for all waves.
- In air columns, the lowest-frequency resonance is called the fundamental, whereas all higher resonant frequencies are called overtones. Collectively, they are called harmonics.

- The resonant frequencies of a tube closed at one end are:

Equation:

$$f_n = n \frac{v_w}{4L}, n = 1, 3, 5...$$

f_1 is the fundamental and L is the length of the tube.

- The resonant frequencies of a tube open at both ends are:

Equation:

$$f_n = n \frac{v_w}{2L}, n = 1, 2, 3...$$

Conceptual Questions

Exercise:

Problem:

How does an unamplified guitar produce sounds so much more intense than those of a plucked string held taut by a simple stick?

Exercise:

Problem:

You are given two wind instruments of identical length. One is open at both ends, whereas the other is closed at one end. Which is able to produce the lowest frequency?

Exercise:

Problem:

What is the difference between an overtone and a harmonic? Are all harmonics overtones? Are all overtones harmonics?

Problems & Exercises

Exercise:**Problem:**

A “showy” custom-built car has two brass horns that are supposed to produce the same frequency but actually emit 263.8 and 264.5 Hz. What beat frequency is produced?

Solution:

0.7 Hz

Exercise:**Problem:**

What beat frequencies will be present: (a) If the musical notes A and C are played together (frequencies of 220 and 264 Hz)? (b) If D and F are played together (frequencies of 297 and 352 Hz)? (c) If all four are played together?

Exercise:**Problem:**

What beat frequencies result if a piano hammer hits three strings that emit frequencies of 127.8, 128.1, and 128.3 Hz?

Solution:

0.3 Hz, 0.2 Hz, 0.5 Hz

Exercise:**Problem:**

A piano tuner hears a beat every 2.00 s when listening to a 264.0-Hz tuning fork and a single piano string. What are the two possible frequencies of the string?

Exercise:

Problem:

(a) What is the fundamental frequency of a 0.672-m-long tube, open at both ends, on a day when the speed of sound is 344 m/s? (b) What is the frequency of its second harmonic?

Solution:

(a) 256 Hz

(b) 512 Hz

Exercise:**Problem:**

If a wind instrument, such as a tuba, has a fundamental frequency of 32.0 Hz, what are its first three overtones? It is closed at one end. (The overtones of a real tuba are more complex than this example, because it is a tapered tube.)

Exercise:**Problem:**

What are the first three overtones of a bassoon that has a fundamental frequency of 90.0 Hz? It is open at both ends. (The overtones of a real bassoon are more complex than this example, because its double reed makes it act more like a tube closed at one end.)

Solution:

180 Hz, 270 Hz, 360 Hz

Exercise:

Problem:

How long must a flute be in order to have a fundamental frequency of 262 Hz (this frequency corresponds to middle C on the evenly tempered chromatic scale) on a day when air temperature is 20.0°C ? It is open at both ends.

Exercise:**Problem:**

What length should an oboe have to produce a fundamental frequency of 110 Hz on a day when the speed of sound is 343 m/s? It is open at both ends.

Solution:

1.56 m

Exercise:**Problem:**

What is the length of a tube that has a fundamental frequency of 176 Hz and a first overtone of 352 Hz if the speed of sound is 343 m/s?

Exercise:**Problem:**

(a) Find the length of an organ pipe closed at one end that produces a fundamental frequency of 256 Hz when air temperature is 18.0°C . (b) What is its fundamental frequency at 25.0°C ?

Solution:

(a) 0.334 m

(b) 259 Hz

Exercise:

Problem:

By what fraction will the frequencies produced by a wind instrument change when air temperature goes from 10.0°C to 30.0°C ? That is, find the ratio of the frequencies at those temperatures.

Exercise:**Problem:**

The ear canal resonates like a tube closed at one end. (See [\[link\]](#).) If ear canals range in length from 1.80 to 2.60 cm in an average population, what is the range of fundamental resonant frequencies? Take air temperature to be 37.0°C , which is the same as body temperature. How does this result correlate with the intensity versus frequency graph ([\[link\]](#)) of the human ear?

Solution:

3.39 to 4.90 kHz

Exercise:**Problem:**

Calculate the first overtone in an ear canal, which resonates like a 2.40-cm-long tube closed at one end, by taking air temperature to be 37.0°C . Is the ear particularly sensitive to such a frequency? (The resonances of the ear canal are complicated by its nonuniform shape, which we shall ignore.)

Exercise:**Problem:**

A crude approximation of voice production is to consider the breathing passages and mouth to be a resonating tube closed at one end. (See [\[link\]](#).) (a) What is the fundamental frequency if the tube is 0.240-m long, by taking air temperature to be 37.0°C ? (b) What would this frequency become if the person replaced the air with helium? Assume the same temperature dependence for helium as for air.

Solution:

(a) 367 Hz

(b) 1.07 kHz

Exercise:**Problem:**

(a) Students in a physics lab are asked to find the length of an air column in a tube closed at one end that has a fundamental frequency of 256 Hz. They hold the tube vertically and fill it with water to the top, then lower the water while a 256-Hz tuning fork is rung and listen for the first resonance. What is the air temperature if the resonance occurs for a length of 0.336 m? (b) At what length will they observe the second resonance (first overtone)?

Exercise:**Problem:**

What frequencies will a 1.80-m-long tube produce in the audible range at 20.0°C if: (a) The tube is closed at one end? (b) It is open at both ends?

Solution:

(a) $f_n = n(47.6 \text{ Hz})$, $n = 1, 3, 5, \dots, 419$

(b) $f_n = n(95.3 \text{ Hz})$, $n = 1, 2, 3, \dots, 210$

Glossary

antinode

point of maximum displacement

node

point of zero displacement

fundamental

the lowest-frequency resonance

overtones

all resonant frequencies higher than the fundamental

harmonics

the term used to refer collectively to the fundamental and its overtones

Hearing

- Define hearing, pitch, loudness, timbre, note, tone, phon, ultrasound, and infrasound.
- Compare loudness to frequency and intensity of a sound.
- Identify structures of the inner ear and explain how they relate to sound perception.



Hearing allows this
vocalist, his band, and his
fans to enjoy music.
(credit: West Point Public
Affairs, Flickr)

The human ear has a tremendous range and sensitivity. It can give us a wealth of simple information—such as pitch, loudness, and direction. And from its input we can detect musical quality and nuances of voiced emotion. How is our hearing related to the physical qualities of sound, and how does the hearing mechanism work?

Hearing is the perception of sound. (Perception is commonly defined to be awareness through the senses, a typically circular definition of higher-level processes in living organisms.) Normal human hearing encompasses frequencies from 20 to 20,000 Hz, an impressive range. Sounds below 20 Hz are called **infrasound**, whereas those above 20,000 Hz are **ultrasound**. Neither is perceived by the ear, although infrasound can sometimes be felt as vibrations. When we do hear low-frequency vibrations, such as the

sounds of a diving board, we hear the individual vibrations only because there are higher-frequency sounds in each. Other animals have hearing ranges different from that of humans. Dogs can hear sounds as high as 30,000 Hz, whereas bats and dolphins can hear up to 100,000-Hz sounds. You may have noticed that dogs respond to the sound of a dog whistle which produces sound out of the range of human hearing. Elephants are known to respond to frequencies below 20 Hz.

The perception of frequency is called **pitch**. Most of us have excellent relative pitch, which means that we can tell whether one sound has a different frequency from another. Typically, we can discriminate between two sounds if their frequencies differ by 0.3% or more. For example, 500.0 and 501.5 Hz are noticeably different. Pitch perception is directly related to frequency and is not greatly affected by other physical quantities such as intensity. Musical **notes** are particular sounds that can be produced by most instruments and in Western music have particular names. Combinations of notes constitute music. Some people can identify musical notes, such as A-sharp, C, or E-flat, just by listening to them. This uncommon ability is called perfect pitch.

The ear is remarkably sensitive to low-intensity sounds. The lowest audible intensity or threshold is about 10^{-12} W/m^2 or 0 dB. Sounds as much as 10^{12} more intense can be briefly tolerated. Very few measuring devices are capable of observations over a range of a trillion. The perception of intensity is called **loudness**. At a given frequency, it is possible to discern differences of about 1 dB, and a change of 3 dB is easily noticed. But loudness is not related to intensity alone. Frequency has a major effect on how loud a sound seems. The ear has its maximum sensitivity to frequencies in the range of 2000 to 5000 Hz, so that sounds in this range are perceived as being louder than, say, those at 500 or 10,000 Hz, even when they all have the same intensity. Sounds near the high- and low-frequency extremes of the hearing range seem even less loud, because the ear is even less sensitive at those frequencies. [\[link\]](#) gives the dependence of certain human hearing perceptions on physical quantities.

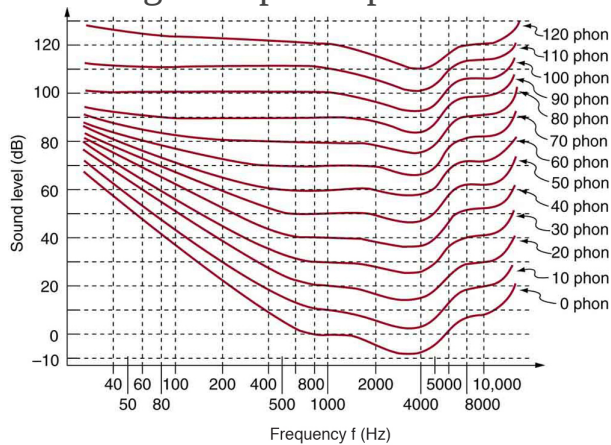
Perception	Physical quantity
Pitch	Frequency
Loudness	Intensity and Frequency
Timbre	Number and relative intensity of multiple frequencies. Subtle craftsmanship leads to non-linear effects and more detail.
Note	Basic unit of music with specific names, combined to generate tunes
Tone	Number and relative intensity of multiple frequencies.

Sound Perceptions

When a violin plays middle C, there is no mistaking it for a piano playing the same note. The reason is that each instrument produces a distinctive set of frequencies and intensities. We call our perception of these combinations of frequencies and intensities **tone** quality, or more commonly the **timbre** of the sound. It is more difficult to correlate timbre perception to physical quantities than it is for loudness or pitch perception. Timbre is more subjective. Terms such as dull, brilliant, warm, cold, pure, and rich are employed to describe the timbre of a sound. So the consideration of timbre takes us into the realm of perceptual psychology, where higher-level processes in the brain are dominant. This is true for other perceptions of sound, such as music and noise. We shall not delve further into them; rather, we will concentrate on the question of loudness perception.

A unit called a **phon** is used to express loudness numerically. Phons differ from decibels because the phon is a unit of loudness perception, whereas the decibel is a unit of physical intensity. [\[link\]](#) shows the relationship of loudness to intensity (or intensity level) and frequency for persons with normal hearing. The curved lines are equal-loudness curves. Each curve is

labeled with its loudness in phons. Any sound along a given curve will be perceived as equally loud by the average person. The curves were determined by having large numbers of people compare the loudness of sounds at different frequencies and sound intensity levels. At a frequency of 1000 Hz, phons are taken to be numerically equal to decibels. The following example helps illustrate how to use the graph:



The relationship of loudness in phons to intensity level (in decibels) and intensity (in watts per meter squared) for persons with normal hearing. The curved lines are equal-loudness curves—all sounds on a given curve are perceived as equally loud. Phons and decibels are defined to be the same at 1000 Hz.

Example:

Measuring Loudness: Loudness Versus Intensity Level and Frequency

(a) What is the loudness in phons of a 100-Hz sound that has an intensity level of 80 dB? (b) What is the intensity level in decibels of a 4000-Hz

sound having a loudness of 70 phons? (c) At what intensity level will an 8000-Hz sound have the same loudness as a 200-Hz sound at 60 dB?

Strategy for (a)

The graph in [\[link\]](#) should be referenced in order to solve this example. To find the loudness of a given sound, you must know its frequency and intensity level and locate that point on the square grid, then interpolate between loudness curves to get the loudness in phons.

Solution for (a)

(1) Identify knowns:

- The square grid of the graph relating phons and decibels is a plot of intensity level versus frequency—both physical quantities.
- 100 Hz at 80 dB lies halfway between the curves marked 70 and 80 phons.

(2) Find the loudness: 75 phons.

Strategy for (b)

The graph in [\[link\]](#) should be referenced in order to solve this example. To find the intensity level of a sound, you must have its frequency and loudness. Once that point is located, the intensity level can be determined from the vertical axis.

Solution for (b)

(1) Identify knowns:

- Values are given to be 4000 Hz at 70 phons.

(2) Follow the 70-phon curve until it reaches 4000 Hz. At that point, it is below the 70 dB line at about 67 dB.

(3) Find the intensity level:

67 dB

Strategy for (c)

The graph in [\[link\]](#) should be referenced in order to solve this example.

Solution for (c)

(1) Locate the point for a 200 Hz and 60 dB sound.

(2) Find the loudness: This point lies just slightly above the 50-phon curve, and so its loudness is 51 phons.

(3) Look for the 51-phon level is at 8000 Hz: 63 dB.

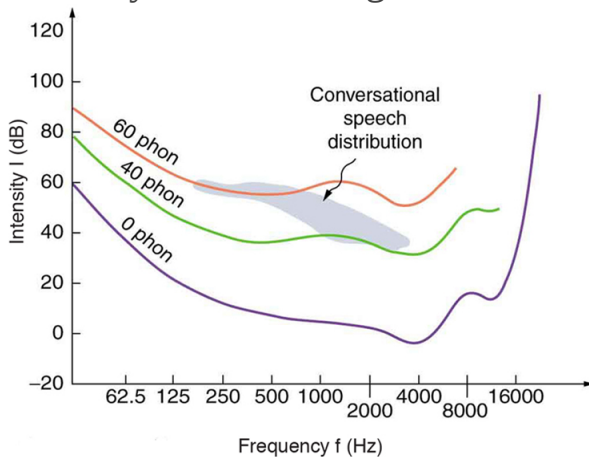
Discussion

These answers, like all information extracted from [\[link\]](#), have uncertainties of several phons or several decibels, partly due to difficulties in interpolation, but mostly related to uncertainties in the equal-loudness curves.

Further examination of the graph in [\[link\]](#) reveals some interesting facts about human hearing. First, sounds below the 0-phon curve are not perceived by most people. So, for example, a 60 Hz sound at 40 dB is inaudible. The 0-phon curve represents the threshold of normal hearing. We can hear some sounds at intensity levels below 0 dB. For example, a 3-dB, 5000-Hz sound is audible, because it lies above the 0-phon curve. The loudness curves all have dips in them between about 2000 and 5000 Hz. These dips mean the ear is most sensitive to frequencies in that range. For example, a 15-dB sound at 4000 Hz has a loudness of 20 phons, the same as a 20-dB sound at 1000 Hz. The curves rise at both extremes of the frequency range, indicating that a greater-intensity level sound is needed at those frequencies to be perceived to be as loud as at middle frequencies. For example, a sound at 10,000 Hz must have an intensity level of 30 dB to seem as loud as a 20 dB sound at 1000 Hz. Sounds above 120 phons are painful as well as damaging.

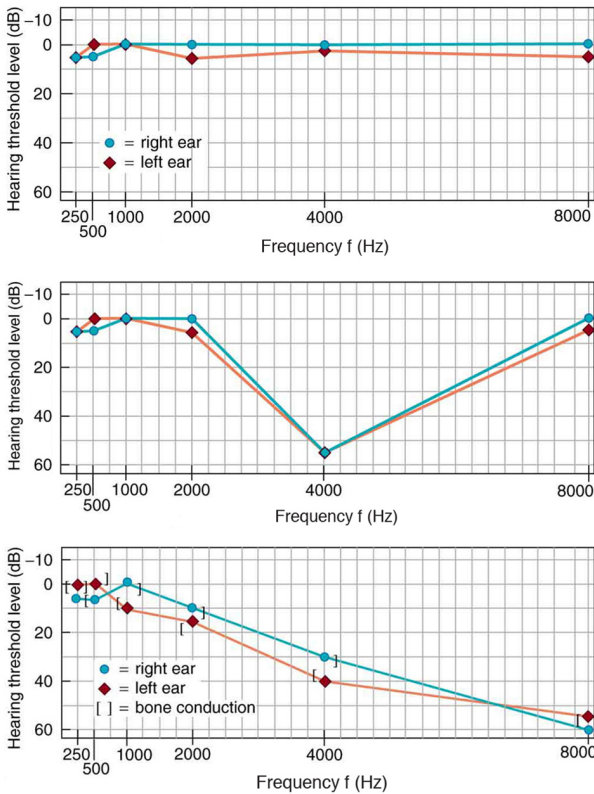
We do not often utilize our full range of hearing. This is particularly true for frequencies above 8000 Hz, which are rare in the environment and are unnecessary for understanding conversation or appreciating music. In fact, people who have lost the ability to hear such high frequencies are usually unaware of their loss until tested. The shaded region in [\[link\]](#) is the frequency and intensity region where most conversational sounds fall. The curved lines indicate what effect hearing losses of 40 and 60 phons will have. A 40-phon hearing loss at all frequencies still allows a person to understand conversation, although it will seem very quiet. A person with a 60-phon loss at all frequencies will hear only the lowest frequencies and will not be able to understand speech unless it is much louder than normal. Even so, speech may seem indistinct, because higher frequencies are not as well perceived. The conversational speech region also has a gender component, in that female voices are usually characterized by higher

frequencies. So the person with a 60-phon hearing impediment might have difficulty understanding the normal conversation of a woman.



The shaded region represents frequencies and intensity levels found in normal conversational speech. The 0-phon line represents the normal hearing threshold, while those at 40 and 60 represent thresholds for people with 40- and 60-phon hearing losses, respectively.

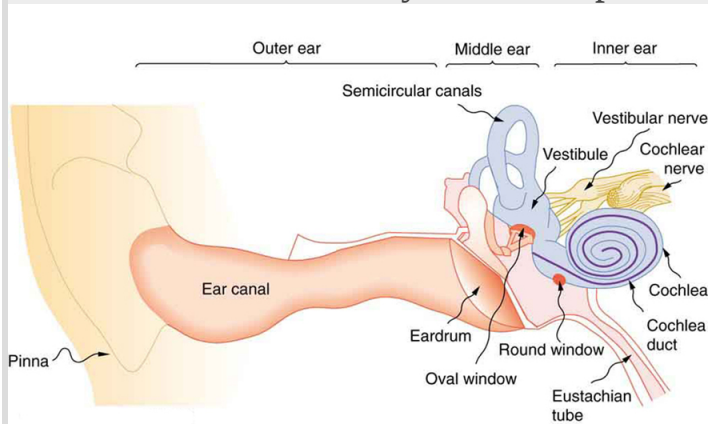
Hearing tests are performed over a range of frequencies, usually from 250 to 8000 Hz, and can be displayed graphically in an audiogram like that in [\[link\]](#). The hearing threshold is measured in dB *relative to the normal threshold*, so that normal hearing registers as 0 dB at all frequencies. Hearing loss caused by noise typically shows a dip near the 4000 Hz frequency, irrespective of the frequency that caused the loss and often affects both ears. The most common form of hearing loss comes with age and is called *presbycusis*—literally elder ear. Such loss is increasingly severe at higher frequencies, and interferes with music appreciation and speech recognition.



Audiograms showing the threshold in intensity level versus frequency for three different individuals. Intensity level is measured relative to the normal threshold. The top left graph is that of a person with normal hearing. The graph to its right has a dip at 4000 Hz and is that of a child who suffered hearing loss due to a cap gun. The third graph is typical of presbycusis, the progressive loss of higher frequency hearing with age. Tests performed by bone conduction (brackets) can distinguish nerve damage from middle ear damage.

Note:**The Hearing Mechanism**

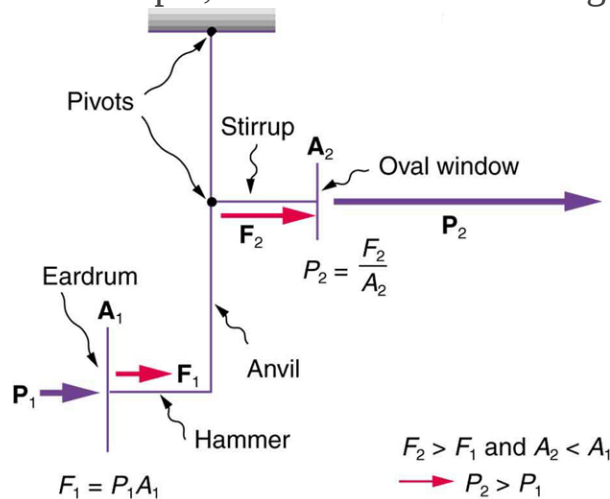
The hearing mechanism involves some interesting physics. The sound wave that impinges upon our ear is a pressure wave. The ear is a transducer that converts sound waves into electrical nerve impulses in a manner much more sophisticated than, but analogous to, a microphone. [\[link\]](#) shows the gross anatomy of the ear with its division into three parts: the outer ear or ear canal; the middle ear, which runs from the eardrum to the cochlea; and the inner ear, which is the cochlea itself. The body part normally referred to as the ear is technically called the pinna.



The illustration shows the gross anatomy of the human ear.

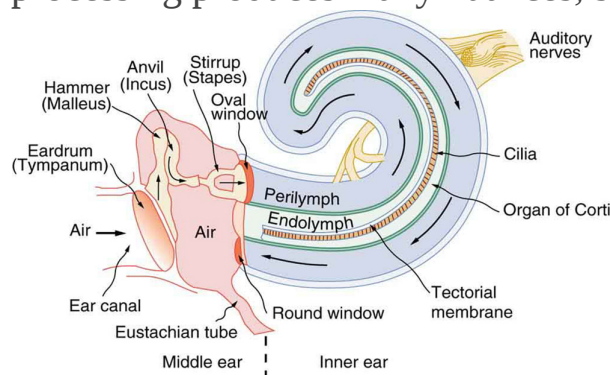
The outer ear, or ear canal, carries sound to the recessed protected eardrum. The air column in the ear canal resonates and is partially responsible for the sensitivity of the ear to sounds in the 2000 to 5000 Hz range. The middle ear converts sound into mechanical vibrations and applies these vibrations to the cochlea. The lever system of the middle ear takes the force exerted on the eardrum by sound pressure variations, amplifies it and transmits it to the

inner ear via the oval window, creating pressure waves in the cochlea approximately 40 times greater than those impinging on the eardrum. (See [\[link\]](#).) Two muscles in the middle ear (not shown) protect the inner ear from very intense sounds. They react to intense sound in a few milliseconds and reduce the force transmitted to the cochlea. This protective reaction can also be triggered by your own voice, so that humming while shooting a gun, for example, can reduce noise damage.



This schematic shows the middle ear's system for converting sound pressure into force, increasing that force through a lever system, and applying the increased force to a small area of the cochlea, thereby creating a pressure about 40 times that in the original sound wave. A protective muscle reaction to intense sounds greatly reduces the mechanical advantage of the lever system.

[\[link\]](#) shows the middle and inner ear in greater detail. Pressure waves moving through the cochlea cause the tectorial membrane to vibrate, rubbing cilia (called hair cells), which stimulate nerves that send electrical signals to the brain. The membrane resonates at different positions for different frequencies, with high frequencies stimulating nerves at the near end and low frequencies at the far end. The complete operation of the cochlea is still not understood, but several mechanisms for sending information to the brain are known to be involved. For sounds below about 1000 Hz, the nerves send signals at the same frequency as the sound. For frequencies greater than about 1000 Hz, the nerves signal frequency by position. There is a structure to the cilia, and there are connections between nerve cells that perform signal processing before information is sent to the brain. Intensity information is partly indicated by the number of nerve signals and by volleys of signals. The brain processes the cochlear nerve signals to provide additional information such as source direction (based on time and intensity comparisons of sounds from both ears). Higher-level processing produces many nuances, such as music appreciation.



The inner ear, or cochlea, is a coiled tube about 3 mm in diameter and 3 cm in length if uncoiled. When the oval window is forced inward, as shown, a pressure wave travels through the perilymph in the direction of the arrows, stimulating nerves at the base of cilia in the organ of Corti.

Hearing losses can occur because of problems in the middle or inner ear. Conductive losses in the middle ear can be partially overcome by sending sound vibrations to the cochlea through the skull. Hearing aids for this purpose usually press against the bone behind the ear, rather than simply amplifying the sound sent into the ear canal as many hearing aids do. Damage to the nerves in the cochlea is not repairable, but amplification can partially compensate. There is a risk that amplification will produce further damage. Another common failure in the cochlea is damage or loss of the cilia but with nerves remaining functional. Cochlear implants that stimulate the nerves directly are now available and widely accepted. Over 100,000 implants are in use, in about equal numbers of adults and children.

The cochlear implant was pioneered in Melbourne, Australia, by Graeme Clark in the 1970s for his deaf father. The implant consists of three external components and two internal components. The external components are a microphone for picking up sound and converting it into an electrical signal, a speech processor to select certain frequencies and a transmitter to transfer the signal to the internal components through electromagnetic induction. The internal components consist of a receiver/transmitter secured in the bone beneath the skin, which converts the signals into electric impulses and sends them through an internal cable to the cochlea and an array of about 24 electrodes wound through the cochlea. These electrodes in turn send the impulses directly into the brain. The electrodes basically emulate the cilia.

Exercise:

Check Your Understanding

Problem:

Are ultrasound and infrasound imperceptible to all hearing organisms?
Explain your answer.

Solution:

No, the range of perceptible sound is based in the range of human hearing. Many other organisms perceive either infrasound or ultrasound.

Section Summary

- The range of audible frequencies is 20 to 20,000 Hz.
- Those sounds above 20,000 Hz are ultrasound, whereas those below 20 Hz are infrasound.
- The perception of frequency is pitch.
- The perception of intensity is loudness.
- Loudness has units of phons.

Conceptual Questions

Exercise:

Problem:

Why can a hearing test show that your threshold of hearing is 0 dB at 250 Hz, when [\[link\]](#) implies that no one can hear such a frequency at less than 20 dB?

Problems & Exercises

Exercise:

Problem:

The factor of 10^{-12} in the range of intensities to which the ear can respond, from threshold to that causing damage after brief exposure, is truly remarkable. If you could measure distances over the same range with a single instrument and the smallest distance you could measure was 1 mm, what would the largest be?

Solution:

Equation:

$$1 \times 10^6 \text{ km}$$

Exercise:

Problem:

The frequencies to which the ear responds vary by a factor of 10^3 . Suppose the speedometer on your car measured speeds differing by the same factor of 10^3 , and the greatest speed it reads is 90.0 mi/h. What would be the slowest nonzero speed it could read?

Exercise:**Problem:**

What are the closest frequencies to 500 Hz that an average person can clearly distinguish as being different in frequency from 500 Hz? The sounds are not present simultaneously.

Solution:

498.5 or 501.5 Hz

Exercise:**Problem:**

Can the average person tell that a 2002-Hz sound has a different frequency than a 1999-Hz sound without playing them simultaneously?

Exercise:**Problem:**

If your radio is producing an average sound intensity level of 85 dB, what is the next lowest sound intensity level that is clearly less intense?

Solution:

82 dB

Exercise:

Problem:

Can you tell that your roommate turned up the sound on the TV if its average sound intensity level goes from 70 to 73 dB?

Exercise:**Problem:**

Based on the graph in [\[link\]](#), what is the threshold of hearing in decibels for frequencies of 60, 400, 1000, 4000, and 15,000 Hz? Note that many AC electrical appliances produce 60 Hz, music is commonly 400 Hz, a reference frequency is 1000 Hz, your maximum sensitivity is near 4000 Hz, and many older TVs produce a 15,750 Hz whine.

Solution:

approximately 48, 9, 0, -7, and 20 dB, respectively

Exercise:**Problem:**

What sound intensity levels must sounds of frequencies 60, 3000, and 8000 Hz have in order to have the same loudness as a 40-dB sound of frequency 1000 Hz (that is, to have a loudness of 40 phons)?

Exercise:**Problem:**

What is the approximate sound intensity level in decibels of a 600-Hz tone if it has a loudness of 20 phons? If it has a loudness of 70 phons?

Solution:

(a) 23 dB

(b) 70 dB

Exercise:

Problem:

(a) What are the loudnesses in phons of sounds having frequencies of 200, 1000, 5000, and 10,000 Hz, if they are all at the same 60.0-dB sound intensity level? (b) If they are all at 110 dB? (c) If they are all at 20.0 dB?

Exercise:**Problem:**

Suppose a person has a 50-dB hearing loss at all frequencies. By how many factors of 10 will low-intensity sounds need to be amplified to seem normal to this person? Note that smaller amplification is appropriate for more intense sounds to avoid further hearing damage.

Solution:

Five factors of 10

Exercise:**Problem:**

If a woman needs an amplification of 5.0×10^{12} times the threshold intensity to enable her to hear at all frequencies, what is her overall hearing loss in dB? Note that smaller amplification is appropriate for more intense sounds to avoid further damage to her hearing from levels above 90 dB.

Exercise:**Problem:**

(a) What is the intensity in watts per meter squared of a just barely audible 200-Hz sound? (b) What is the intensity in watts per meter squared of a barely audible 4000-Hz sound?

Solution:

(a) $2 \times 10^{-10} \text{ W/m}^2$

(b) $2 \times 10^{-13} \text{ W/m}^2$

Exercise:

Problem:

(a) Find the intensity in watts per meter squared of a 60.0-Hz sound having a loudness of 60 phons. (b) Find the intensity in watts per meter squared of a 10,000-Hz sound having a loudness of 60 phons.

Exercise:

Problem:

A person has a hearing threshold 10 dB above normal at 100 Hz and 50 dB above normal at 4000 Hz. How much more intense must a 100-Hz tone be than a 4000-Hz tone if they are both barely audible to this person?

Solution:

2.5

Exercise:

Problem:

A child has a hearing loss of 60 dB near 5000 Hz, due to noise exposure, and normal hearing elsewhere. How much more intense is a 5000-Hz tone than a 400-Hz tone if they are both barely audible to the child?

Exercise:

Problem:

What is the ratio of intensities of two sounds of identical frequency if the first is just barely discernible as louder to a person than the second?

Solution:

1.26

Glossary

loudness

the perception of sound intensity

timbre

number and relative intensity of multiple sound frequencies

note

basic unit of music with specific names, combined to generate tunes

tone

number and relative intensity of multiple sound frequencies

phon

the numerical unit of loudness

ultrasound

sounds above 20,000 Hz

infrasound

sounds below 20 Hz

Ultrasound

- Define acoustic impedance and intensity reflection coefficient.
- Describe medical and other uses of ultrasound technology.
- Calculate acoustic impedance using density values and the speed of ultrasound.
- Calculate the velocity of a moving object using Doppler-shifted ultrasound.



Ultrasound is used in medicine to painlessly and noninvasively monitor patient health and diagnose a wide range of disorders. (credit: abbybatchelder, Flickr)

Any sound with a frequency above 20,000 Hz (or 20 kHz)—that is, above the highest audible frequency—is defined to be ultrasound. In practice, it is possible to create ultrasound frequencies up to more than a gigahertz. (Higher frequencies are difficult to create; furthermore, they propagate poorly because they are very strongly absorbed.) Ultrasound has a tremendous number of applications, which range from burglar alarms to use in cleaning delicate objects to the guidance systems of bats. We begin our discussion of ultrasound with some of its applications in medicine, in which it is used extensively both for diagnosis and for therapy.

Note:

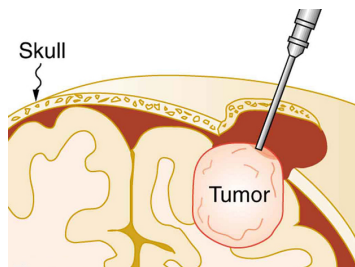
Characteristics of Ultrasound

The characteristics of ultrasound, such as frequency and intensity, are wave properties common to all types of waves. Ultrasound also has a wavelength that limits the fineness of detail it can detect. This characteristic is true of all waves. We can never observe details significantly smaller than the wavelength of our probe; for example,

we will never see individual atoms with visible light, because the atoms are so small compared with the wavelength of light.

Ultrasound in Medical Therapy

Ultrasound, like any wave, carries energy that can be absorbed by the medium carrying it, producing effects that vary with intensity. When focused to intensities of 10^3 to 10^5 W/m², ultrasound can be used to shatter gallstones or pulverize cancerous tissue in surgical procedures. (See [\[link\]](#).) Intensities this great can damage individual cells, variously causing their protoplasm to stream inside them, altering their permeability, or rupturing their walls through *cavitation*. Cavitation is the creation of vapor cavities in a fluid—the longitudinal vibrations in ultrasound alternatively compress and expand the medium, and at sufficient amplitudes the expansion separates molecules. Most cavitation damage is done when the cavities collapse, producing even greater shock pressures.



The tip of this small probe oscillates at 23 kHz with such a large amplitude that it pulverizes tissue on contact. The debris is then aspirated. The speed of the tip may exceed the speed of sound in tissue, thus creating shock waves and cavitation, rather than a smooth

simple harmonic
oscillator-type
wave.

Most of the energy carried by high-intensity ultrasound in tissue is converted to thermal energy. In fact, intensities of 10^3 to 10^4 W/m² are commonly used for deep-heat treatments called ultrasound diathermy. Frequencies of 0.8 to 1 MHz are typical. In both athletics and physical therapy, ultrasound diathermy is most often applied to injured or overworked muscles to relieve pain and improve flexibility. Skill is needed by the therapist to avoid “bone burns” and other tissue damage caused by overheating and cavitation, sometimes made worse by reflection and focusing of the ultrasound by joint and bone tissue.

In some instances, you may encounter a different decibel scale, called the sound *pressure* level, when ultrasound travels in water or in human and other biological tissues. We shall not use the scale here, but it is notable that numbers for sound pressure levels range 60 to 70 dB higher than you would quote for β , the sound intensity level used in this text. Should you encounter a sound pressure level of 220 decibels, then, it is not an astronomically high intensity, but equivalent to about 155 dB—high enough to destroy tissue, but not as unreasonably high as it might seem at first.

Ultrasound in Medical Diagnostics

When used for imaging, ultrasonic waves are emitted from a transducer, a crystal exhibiting the piezoelectric effect (the expansion and contraction of a substance when a voltage is applied across it, causing a vibration of the crystal). These high-frequency vibrations are transmitted into any tissue in contact with the transducer. Similarly, if a pressure is applied to the crystal (in the form of a wave reflected off tissue layers), a voltage is produced which can be recorded. The crystal therefore acts as both a transmitter and a receiver of sound. Ultrasound is also partially absorbed by tissue on its path, both on its journey away from the transducer and on its return journey. From the time between when the original signal is sent and when the reflections from various boundaries between media are received, (as well as a measure of the intensity loss of the signal), the nature and position of each boundary between tissues and organs may be deduced.

Reflections at boundaries between two different media occur because of differences in a characteristic known as the **acoustic impedance** Z of each substance. Impedance is defined as

Equation:

$$Z = \rho v,$$

where ρ is the density of the medium (in kg/m^3) and v is the speed of sound through the medium (in m/s). The units for Z are therefore $\text{kg}/(\text{m}^2 \cdot \text{s})$.

[\[link\]](#) shows the density and speed of sound through various media (including various soft tissues) and the associated acoustic impedances. Note that the acoustic impedances for soft tissue do not vary much but that there is a big difference between the acoustic impedance of soft tissue and air and also between soft tissue and bone.

Medium	Density (kg/m^3)	Speed of Ultrasound (m/s)	Acoustic Impedance ($\text{kg}/(\text{m}^2 \cdot \text{s})$)
Air	1.3	330	429
Water	1000	1500	1.5×10^6
Blood	1060	1570	1.66×10^6
Fat	925	1450	1.34×10^6
Muscle (average)	1075	1590	1.70×10^6
Bone (varies)	1400– 1900	4080	5.7×10^6 to 7.8×10^6
Barium titanate (transducer material)	5600	5500	30.8×10^6

The Ultrasound Properties of Various Media, Including Soft Tissue Found in the Body

At the boundary between media of different acoustic impedances, some of the wave energy is reflected and some is transmitted. The greater the *difference* in acoustic impedance between the two media, the greater the reflection and the smaller the transmission.

The **intensity reflection coefficient** a is defined as the ratio of the intensity of the reflected wave relative to the incident (transmitted) wave. This statement can be written mathematically as

Equation:

$$a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2},$$

where Z_1 and Z_2 are the acoustic impedances of the two media making up the boundary. A reflection coefficient of zero (corresponding to total transmission and no reflection) occurs when the acoustic impedances of the two media are the same. An impedance “match” (no reflection) provides an efficient coupling of sound energy from one medium to another. The image formed in an ultrasound is made by tracking reflections (as shown in [\[link\]](#)) and mapping the intensity of the reflected sound waves in a two-dimensional plane.

Example:

Calculate Acoustic Impedance and Intensity Reflection Coefficient: Ultrasound and Fat Tissue

(a) Using the values for density and the speed of ultrasound given in [\[link\]](#), show that the acoustic impedance of fat tissue is indeed $1.34 \times 10^6 \text{ kg}/(\text{m}^2 \cdot \text{s})$.

(b) Calculate the intensity reflection coefficient of ultrasound when going from fat to muscle tissue.

Strategy for (a)

The acoustic impedance can be calculated using $Z = \rho v$ and the values for ρ and v found in [\[link\]](#).

Solution for (a)

(1) Substitute known values from [\[link\]](#) into $Z = \rho v$.

Equation:

$$Z = \rho v = (925 \text{ kg}/\text{m}^3)(1450 \text{ m/s})$$

(2) Calculate to find the acoustic impedance of fat tissue.

Equation:

$$1.34 \times 10^6 \text{ kg}/(\text{m}^2 \cdot \text{s})$$

This value is the same as the value given for the acoustic impedance of fat tissue.

Strategy for (b)

The intensity reflection coefficient for any boundary between two media is given by

$$a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}, \text{ and the acoustic impedance of muscle is given in [\[link\]](#)}.$$

Solution for (b)

Substitute known values into $a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}$ to find the intensity reflection coefficient:

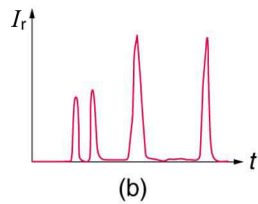
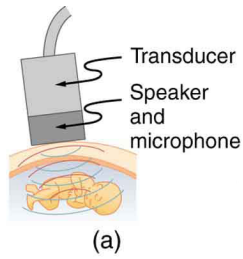
Equation:

$$a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2} = \frac{\left(1.34 \times 10^6 \text{ kg}/(\text{m}^2 \cdot \text{s}) - 1.70 \times 10^6 \text{ kg}/(\text{m}^2 \cdot \text{s})\right)^2}{\left(1.70 \times 10^6 \text{ kg}/(\text{m}^2 \cdot \text{s}) + 1.34 \times 10^6 \text{ kg}/(\text{m}^2 \cdot \text{s})\right)^2} = 0.014$$

Discussion

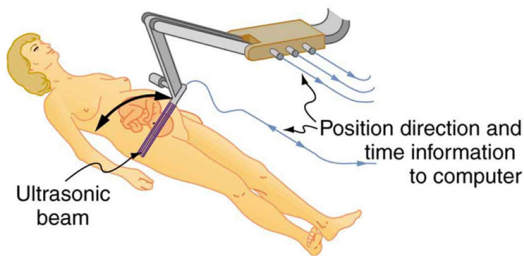
This result means that only 1.4% of the incident intensity is reflected, with the remaining being transmitted.

The applications of ultrasound in medical diagnostics have produced untold benefits with no known risks. Diagnostic intensities are too low (about $10^{-2} \text{ W}/\text{m}^2$) to cause thermal damage. More significantly, ultrasound has been in use for several decades and detailed follow-up studies do not show evidence of ill effects, quite unlike the case for x-rays.



(a) An ultrasound speaker doubles as a microphone. Brief bleeps are broadcast, and echoes are recorded from various depths. (b) Graph of echo intensity versus time. The time for echoes to return is directly proportional to the distance of the reflector, yielding this information noninvasively.

The most common ultrasound applications produce an image like that shown in [\[link\]](#). The speaker-microphone broadcasts a directional beam, sweeping the beam across the area of interest. This is accomplished by having multiple ultrasound sources in the probe's head, which are phased to interfere constructively in a given, adjustable direction. Echoes are measured as a function of position as well as depth. A computer constructs an image that reveals the shape and density of internal structures.



(a)



(b)

(a) An ultrasonic image is produced by sweeping the ultrasonic beam across the area of interest, in this case the woman's abdomen. Data are recorded and analyzed in a computer, providing a two-dimensional image. (b) Ultrasound image of 12-week-old fetus. (credit: Margaret W. Carruthers, Flickr)

How much detail can ultrasound reveal? The image in [\[link\]](#) is typical of low-cost systems, but that in [\[link\]](#) shows the remarkable detail possible with more advanced systems, including 3D imaging. Ultrasound today is commonly used in prenatal care. Such imaging can be used to see if the fetus is developing at a normal rate, and help in the determination of serious problems early in the pregnancy. Ultrasound is also in wide use to image the chambers of the heart and the flow of blood within the beating heart, using the Doppler effect (echocardiology).

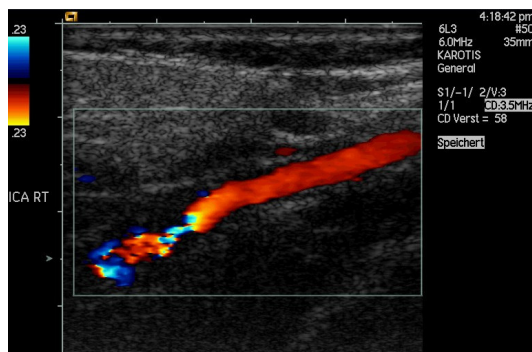
Whenever a wave is used as a probe, it is very difficult to detect details smaller than its wavelength λ . Indeed, current technology cannot do quite this well. Abdominal scans may use a 7-MHz frequency, and the speed of sound in tissue is about 1540 m/s—so the wavelength limit to detail would be $\lambda = \frac{v_w}{f} = \frac{1540 \text{ m/s}}{7 \times 10^6 \text{ Hz}} = 0.22 \text{ mm}$. In practice, 1-mm detail is attainable, which is sufficient for many purposes. Higher-frequency ultrasound would allow greater detail, but it does not penetrate as well as lower frequencies do. The accepted rule of thumb is that you can effectively scan to a depth of about 500λ into tissue. For 7 MHz, this penetration limit is $500 \times 0.22 \text{ mm}$, which is 0.11 m. Higher frequencies may be employed in smaller organs, such as the eye, but are not practical for looking deep into the body.



A 3D ultrasound image of a fetus. As well as for the detection of any abnormalities, such scans have also been shown to be useful for strengthening the emotional bonding between parents and their unborn child. (credit: Jennie Cu, Wikimedia Commons)

In addition to shape information, ultrasonic scans can produce density information superior to that found in X-rays, because the intensity of a reflected sound is related to changes in density. Sound is most strongly reflected at places where density changes are greatest.

Another major use of ultrasound in medical diagnostics is to detect motion and determine velocity through the Doppler shift of an echo, known as **Doppler-shifted ultrasound**. This technique is used to monitor fetal heartbeat, measure blood velocity, and detect occlusions in blood vessels, for example. (See [\[link\]](#).) The magnitude of the Doppler shift in an echo is directly proportional to the velocity of whatever reflects the sound. Because an echo is involved, there is actually a double shift. The first occurs because the reflector (say a fetal heart) is a moving observer and receives a Doppler-shifted frequency. The reflector then acts as a moving source, producing a second Doppler shift.



This Doppler-shifted ultrasonic image of a partially occluded artery uses color to indicate velocity. The highest velocities are in red, while the lowest are blue. The blood must move faster through the constriction to carry the same flow. (credit: Arning C, Grzyska U, Wikimedia Commons)

A clever technique is used to measure the Doppler shift in an echo. The frequency of the echoed sound is superimposed on the broadcast frequency, producing beats. The beat frequency is $F_B = |f_1 - f_2|$, and so it is directly proportional to the Doppler shift ($f_1 - f_2$) and hence, the reflector's velocity. The advantage in this technique is that the Doppler shift is small (because the reflector's velocity is small), so that great accuracy would be needed to measure the shift directly. But measuring the beat frequency is easy, and it is not affected if the broadcast frequency varies somewhat. Furthermore, the beat frequency is in the audible range and can be amplified for audio feedback to the medical observer.

Note:

Uses for Doppler-Shifted Radar

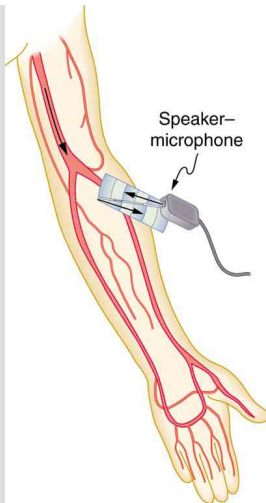
Doppler-shifted radar echoes are used to measure wind velocities in storms as well as aircraft and automobile speeds. The principle is the same as for Doppler-shifted ultrasound. There is evidence that bats and dolphins may also sense the velocity of an object (such as prey) reflecting their ultrasound signals by observing its Doppler shift.

Example:

Calculate Velocity of Blood: Doppler-Shifted Ultrasound

Ultrasound that has a frequency of 2.50 MHz is sent toward blood in an artery that is moving toward the source at 20.0 cm/s, as illustrated in [\[link\]](#). Use the speed of sound in human tissue as 1540 m/s. (Assume that the frequency of 2.50 MHz is accurate to seven significant figures.)

- What frequency does the blood receive?
- What frequency returns to the source?
- What beat frequency is produced if the source and returning frequencies are mixed?



Ultrasound is partly reflected by blood cells and plasma back toward the speaker-microphone. Because the cells are moving, two Doppler shifts are produced—one for blood as a moving observer, and the other for the reflected sound coming from a moving source. The magnitude of the shift is directly proportional to blood velocity.

Strategy

The first two questions can be answered using $f_{\text{obs}} = f_s \left(\frac{v_w}{v_w \pm v_s} \right)$ and

$f_{\text{obs}} = f_s \left(\frac{v_w \pm v_{\text{obs}}}{v_w} \right)$ for the Doppler shift. The last question asks for beat frequency, which is the difference between the original and returning frequencies.

Solution for (a)

(1) Identify knowns:

- The blood is a moving observer, and so the frequency it receives is given by

Equation:

$$f_{\text{obs}} = f_s \left(\frac{v_w \pm v_{\text{obs}}}{v_w} \right).$$

- v_b is the blood velocity (v_{obs} here) and the plus sign is chosen because the motion is toward the source.

(2) Enter the given values into the equation.

Equation:

$$f_{\text{obs}} = (2,500,000 \text{ Hz}) \left(\frac{1540 \text{ m/s} + 0.2 \text{ m/s}}{1540 \text{ m/s}} \right)$$

(3) Calculate to find the frequency: 2,500,325 Hz.

Solution for (b)

(1) Identify knowns:

- The blood acts as a moving source.
- The microphone acts as a stationary observer.
- The frequency leaving the blood is 2,500,325 Hz, but it is shifted upward as given by

Equation:

$$f_{\text{obs}} = f_s \left(\frac{v_w}{v_w - v_b} \right).$$

f_{obs} is the frequency received by the speaker-microphone.

- The source velocity is v_b .
- The minus sign is used because the motion is toward the observer.

The minus sign is used because the motion is toward the observer.

(2) Enter the given values into the equation:

Equation:

$$f_{\text{obs}} = (2,500,325 \text{ Hz}) \left(\frac{1540 \text{ m/s}}{1540 \text{ m/s} - 0.200 \text{ m/s}} \right)$$

(3) Calculate to find the frequency returning to the source: 2,500,649 Hz.

Solution for (c)

(1) Identify knowns:

- The beat frequency is simply the absolute value of the difference between f_s and f_{obs} , as stated in:

Equation:

$$f_B = | f_{\text{obs}} - f_s |.$$

(2) Substitute known values:

Equation:

$$| 2,500,649 \text{ Hz} - 2,500,000 \text{ Hz} |$$

(3) Calculate to find the beat frequency: 649 Hz.

Discussion

The Doppler shifts are quite small compared with the original frequency of 2.50 MHz. It is far easier to measure the beat frequency than it is to measure the echo frequency with an accuracy great enough to see shifts of a few hundred hertz out of a couple of megahertz. Furthermore, variations in the source frequency do not greatly affect the beat frequency, because both f_s and f_{obs} would increase or decrease. Those changes subtract out in $f_B = | f_{\text{obs}} - f_s |$.

Note:

Industrial and Other Applications of Ultrasound

Industrial, retail, and research applications of ultrasound are common. A few are discussed here. Ultrasonic cleaners have many uses. Jewelry, machined parts, and other objects that have odd shapes and crevices are immersed in a cleaning fluid that is agitated with ultrasound typically about 40 kHz in frequency. The intensity is great enough to cause cavitation, which is responsible for most of the cleansing action. Because cavitation-produced shock pressures are large and well transmitted in a fluid,

they reach into small crevices where even a low-surface-tension cleaning fluid might not penetrate.

Sonar is a familiar application of ultrasound. Sonar typically employs ultrasonic frequencies in the range from 30.0 to 100 kHz. Bats, dolphins, submarines, and even some birds use ultrasonic sonar. Echoes are analyzed to give distance and size information both for guidance and finding prey. In most sonar applications, the sound reflects quite well because the objects of interest have significantly different density than the medium in which they travel. When the Doppler shift is observed, velocity information can also be obtained. Submarine sonar can be used to obtain such information, and there is evidence that some bats also sense velocity from their echoes.

Similarly, there are a range of relatively inexpensive devices that measure distance by timing ultrasonic echoes. Many cameras, for example, use such information to focus automatically. Some doors open when their ultrasonic ranging devices detect a nearby object, and certain home security lights turn on when their ultrasonic rangefinders observe motion. Ultrasonic “measuring tapes” also exist to measure such things as room dimensions. Sinks in public restrooms are sometimes automated with ultrasound devices to turn faucets on and off when people wash their hands. These devices reduce the spread of germs and can conserve water.

Ultrasound is used for nondestructive testing in industry and by the military. Because ultrasound reflects well from any large change in density, it can reveal cracks and voids in solids, such as aircraft wings, that are too small to be seen with x-rays. For similar reasons, ultrasound is also good for measuring the thickness of coatings, particularly where there are several layers involved.

Basic research in solid state physics employs ultrasound. Its attenuation is related to a number of physical characteristics, making it a useful probe. Among these characteristics are structural changes such as those found in liquid crystals, the transition of a material to a superconducting phase, as well as density and other properties.

These examples of the uses of ultrasound are meant to whet the appetites of the curious, as well as to illustrate the underlying physics of ultrasound. There are many more applications, as you can easily discover for yourself.

Exercise:

Check Your Understanding

Problem:

Why is it possible to use ultrasound both to observe a fetus in the womb and also to destroy cancerous tumors in the body?

Solution:

Ultrasound can be used medically at different intensities. Lower intensities do not cause damage and are used for medical imaging. Higher intensities can pulverize and destroy targeted substances in the body, such as tumors.

Section Summary

- The acoustic impedance is defined as:

Equation:

$$Z = \rho v,$$

ρ is the density of a medium through which the sound travels and v is the speed of sound through that medium.

- The intensity reflection coefficient a , a measure of the ratio of the intensity of the wave reflected off a boundary between two media relative to the intensity of the incident wave, is given by

Equation:

$$a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}.$$

- The intensity reflection coefficient is a unitless quantity.

Conceptual Questions**Exercise:****Problem:**

If audible sound follows a rule of thumb similar to that for ultrasound, in terms of its absorption, would you expect the high or low frequencies from your neighbor's stereo to penetrate into your house? How does this expectation compare with your experience?

Exercise:**Problem:**

Elephants and whales are known to use infrasound to communicate over very large distances. What are the advantages of infrasound for long distance communication?

Exercise:**Problem:**

It is more difficult to obtain a high-resolution ultrasound image in the abdominal region of someone who is overweight than for someone who has a slight build. Explain why this statement is accurate.

Exercise:**Problem:**

Suppose you read that 210-dB ultrasound is being used to pulverize cancerous tumors. You calculate the intensity in watts per centimeter squared and find it is unreasonably high (10^5 W/cm^2). What is a possible explanation?

Problems & Exercises

Unless otherwise indicated, for problems in this section, assume that the speed of sound through human tissues is 1540 m/s.

Exercise:**Problem:**

What is the sound intensity level in decibels of ultrasound of intensity 10^5 W/m^2 , used to pulverize tissue during surgery?

Solution:

170 dB

Exercise:**Problem:**

Is 155-dB ultrasound in the range of intensities used for deep heating? Calculate the intensity of this ultrasound and compare this intensity with values quoted in the text.

Exercise:**Problem:**

Find the sound intensity level in decibels of $2.00 \times 10^{-2} \text{ W/m}^2$ ultrasound used in medical diagnostics.

Solution:

103 dB

Exercise:**Problem:**

The time delay between transmission and the arrival of the reflected wave of a signal using ultrasound traveling through a piece of fat tissue was 0.13 ms. At what depth did this reflection occur?

Exercise:**Problem:**

In the clinical use of ultrasound, transducers are always coupled to the skin by a thin layer of gel or oil, replacing the air that would otherwise exist between the transducer and the skin. (a) Using the values of acoustic impedance given in [\[link\]](#) calculate the intensity reflection coefficient between transducer material and air. (b) Calculate the intensity reflection coefficient between transducer material and gel (assuming for this problem that its acoustic impedance is identical to that of water). (c) Based on the results of your calculations, explain why the gel is used.

Solution:

(a) 1.00

(b) 0.823

(c) Gel is used to facilitate the transmission of the ultrasound between the transducer and the patient's body.

Exercise:**Problem:**

(a) Calculate the minimum frequency of ultrasound that will allow you to see details as small as 0.250 mm in human tissue. (b) What is the effective depth to which this sound is effective as a diagnostic probe?

Exercise:

Problem:

(a) Find the size of the smallest detail observable in human tissue with 20.0-MHz ultrasound. (b) Is its effective penetration depth great enough to examine the entire eye (about 3.00 cm is needed)? (c) What is the wavelength of such ultrasound in 0°C air?

Solution:

(a) 77.0 μm

(b) Effective penetration depth = 3.85 cm, which is enough to examine the eye.

(c) 16.6 μm

Exercise:**Problem:**

(a) Echo times are measured by diagnostic ultrasound scanners to determine distances to reflecting surfaces in a patient. What is the difference in echo times for tissues that are 3.50 and 3.60 cm beneath the surface? (This difference is the minimum resolving time for the scanner to see details as small as 0.100 cm, or 1.00 mm. Discrimination of smaller time differences is needed to see smaller details.) (b) Discuss whether the period T of this ultrasound must be smaller than the minimum time resolution. If so, what is the minimum frequency of the ultrasound and is that out of the normal range for diagnostic ultrasound?

Exercise:**Problem:**

(a) How far apart are two layers of tissue that produce echoes having round-trip times (used to measure distances) that differ by 0.750 μs ? (b) What minimum frequency must the ultrasound have to see detail this small?

Solution:

(a) $5.78 \times 10^{-4} \text{ m}$

(b) $2.67 \times 10^6 \text{ Hz}$

Exercise:

Problem:

(a) A bat uses ultrasound to find its way among trees. If this bat can detect echoes 1.00 ms apart, what minimum distance between objects can it detect? (b) Could this distance explain the difficulty that bats have finding an open door when they accidentally get into a house?

Exercise:**Problem:**

A dolphin is able to tell in the dark that the ultrasound echoes received from two sharks come from two different objects only if the sharks are separated by 3.50 m, one being that much farther away than the other. (a) If the ultrasound has a frequency of 100 kHz, show this ability is not limited by its wavelength. (b) If this ability is due to the dolphin's ability to detect the arrival times of echoes, what is the minimum time difference the dolphin can perceive?

Solution:

(a) $v_w = 1540 \text{ m/s} = f\lambda \Rightarrow \lambda = \frac{1540 \text{ m/s}}{100 \times 10^3 \text{ Hz}} = 0.0154 \text{ m} < 3.50 \text{ m}$. Because the wavelength is much shorter than the distance in question, the wavelength is not the limiting factor.

(b) 4.55 ms

Exercise:**Problem:**

A diagnostic ultrasound echo is reflected from moving blood and returns with a frequency 500 Hz higher than its original 2.00 MHz. What is the velocity of the blood? (Assume that the frequency of 2.00 MHz is accurate to seven significant figures and 500 Hz is accurate to three significant figures.)

Exercise:**Problem:**

Ultrasound reflected from an oncoming bloodstream that is moving at 30.0 cm/s is mixed with the original frequency of 2.50 MHz to produce beats. What is the beat frequency? (Assume that the frequency of 2.50 MHz is accurate to seven significant figures.)

Solution:

974 Hz

(Note: extra digits were retained in order to show the difference.)

Glossary

acoustic impedance

property of medium that makes the propagation of sound waves more difficult

intensity reflection coefficient

a measure of the ratio of the intensity of the wave reflected off a boundary between two media relative to the intensity of the incident wave

Doppler-shifted ultrasound

a medical technique to detect motion and determine velocity through the Doppler shift of an echo

Introduction to Geometric Optics

class="introduction"

Geometric Optics

Light from this page or screen is formed into an image by the lens of your eye, much as the lens of the camera that made this photograph. Mirrors, like lenses, can also form images that in turn are captured by your eye.

Image
seen as a
result of
reflection
of light
on a
plane
smooth
surface.
(credit:
NASA
Goddard
Photo
and
Video,
via
Flickr)



Our lives are filled with light. Through vision, the most valued of our senses, light can evoke spiritual emotions, such as when we view a magnificent sunset or glimpse a rainbow breaking through the clouds. Light can also simply amuse us in a theater, or warn us to stop at an intersection. It has innumerable uses beyond vision. Light can carry telephone signals through glass fibers or cook a meal in a solar oven. Life itself could not exist without light's energy. From photosynthesis in plants to the sun warming a cold-blooded animal, its supply of energy is vital.



Double Rainbow over the bay

of Pocitos in Montevideo,
Uruguay. (credit: Madrax,
Wikimedia Commons)

We already know that visible light is the type of electromagnetic waves to which our eyes respond. That knowledge still leaves many questions regarding the nature of light and vision. What is color, and how do our eyes detect it? Why do diamonds sparkle? How does light travel? How do lenses and mirrors form images? These are but a few of the questions that are answered by the study of optics. Optics is the branch of physics that deals with the behavior of visible light and other electromagnetic waves. In particular, optics is concerned with the generation and propagation of light and its interaction with matter. What we have already learned about the generation of light in our study of heat transfer by radiation will be expanded upon in later topics, especially those on atomic physics. Now, we will concentrate on the propagation of light and its interaction with matter.

It is convenient to divide optics into two major parts based on the size of objects that light encounters. When light interacts with an object that is several times as large as the light's wavelength, its observable behavior is like that of a ray; it does not prominently display its wave characteristics. We call this part of optics "geometric optics." This chapter will concentrate on such situations. When light interacts with smaller objects, it has very prominent wave characteristics, such as constructive and destructive interference. [Wave Optics](#) will concentrate on such situations.

The Ray Aspect of Light

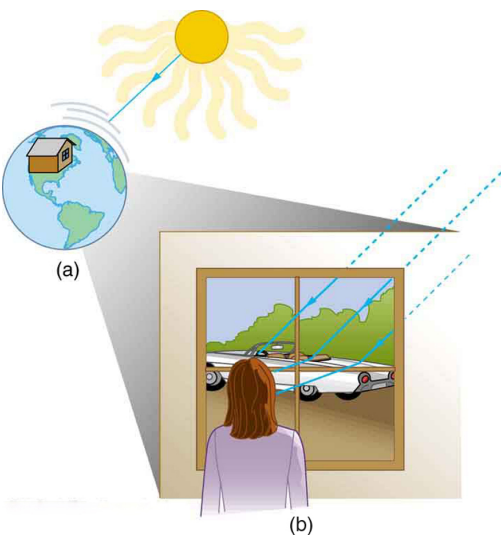
- List the ways by which light travels from a source to another location.

There are three ways in which light can travel from a source to another location. (See [\[link\]](#).) It can come directly from the source through empty space, such as from the Sun to Earth. Or light can travel through various media, such as air and glass, to the person. Light can also arrive after being reflected, such as by a mirror. In all of these cases, light is modeled as traveling in straight lines called rays. Light may change direction when it encounters objects (such as a mirror) or in passing from one material to another (such as in passing from air to glass), but it then continues in a straight line or as a ray. The word **ray** comes from mathematics and here means a straight line that originates at some point. It is acceptable to visualize light rays as laser rays (or even science fiction depictions of ray guns).

Note:

Ray

The word “ray” comes from mathematics and here means a straight line that originates at some point.



Three methods for light to travel from a source to another location. (a) Light reaches the upper atmosphere of Earth traveling through empty space directly from the source. (b) Light can reach a person in one of two ways. It can travel through media like air and glass. It can also reflect from an object like a mirror. In the situations shown here, light interacts with objects large enough that it travels in straight lines, like a ray.

Experiments, as well as our own experiences, show that when light interacts with objects several times as large as its wavelength, it travels in straight lines and acts like a ray. Its wave characteristics are not pronounced in such situations. Since the wavelength of light is less than a micron (a thousandth of a millimeter), it acts like a ray in the many common situations in which it encounters objects larger than a micron. For example, when light encounters anything we can observe with unaided eyes, such as a mirror, it acts like a ray, with only subtle wave characteristics. We will concentrate on the ray characteristics in this chapter.

Since light moves in straight lines, changing directions when it interacts with materials, it is described by geometry and simple trigonometry. This part of optics, where the ray aspect of light dominates, is therefore called **geometric optics**. There are two laws that govern how light changes direction when it interacts with matter. These are the law of reflection, for

situations in which light bounces off matter, and the law of refraction, for situations in which light passes through matter.

Note:

Geometric Optics

The part of optics dealing with the ray aspect of light is called geometric optics.

Section Summary

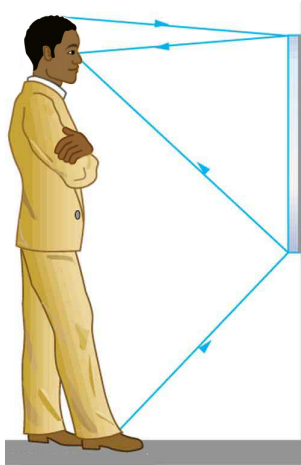
- A straight line that originates at some point is called a ray.
- The part of optics dealing with the ray aspect of light is called geometric optics.
- Light can travel in three ways from a source to another location: (1) directly from the source through empty space; (2) through various media; (3) after being reflected from a mirror.

Problems & Exercises

Exercise:

Problem:

Suppose a man stands in front of a mirror as shown in [\[link\]](#). His eyes are 1.65 m above the floor, and the top of his head is 0.13 m higher. Find the height above the floor of the top and bottom of the smallest mirror in which he can see both the top of his head and his feet. How is this distance related to the man's height?



A full-length mirror is one in which you can see all of yourself. It need not be as big as you, and its size is independent of your distance from it.

Solution:

Top from floor, bottom from floor. Height of mirror is , or precisely one-half the height of the person.

Glossary

ray
straight line that originates at some point

geometric optics

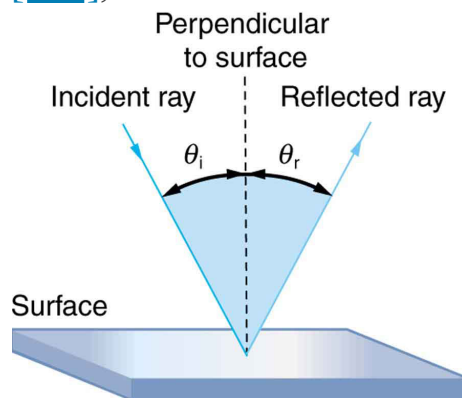
part of optics dealing with the ray aspect of light

The Law of Reflection

- Explain reflection of light from polished and rough surfaces.

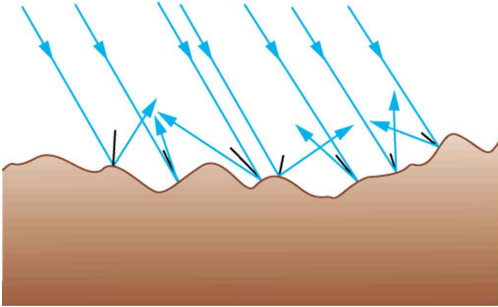
Whenever we look into a mirror, or squint at sunlight glinting from a lake, we are seeing a reflection. When you look at this page, too, you are seeing light reflected from it. Large telescopes use reflection to form an image of stars and other astronomical objects.

The law of reflection is illustrated in [\[link\]](#), which also shows how the angles are measured relative to the perpendicular to the surface at the point where the light ray strikes. We expect to see reflections from smooth surfaces, but [\[link\]](#) illustrates how a rough surface reflects light. Since the light strikes different parts of the surface at different angles, it is reflected in many different directions, or diffused. Diffused light is what allows us to see a sheet of paper from any angle, as illustrated in [\[link\]](#). Many objects, such as people, clothing, leaves, and walls, have rough surfaces and can be seen from all sides. A mirror, on the other hand, has a smooth surface (compared with the wavelength of light) and reflects light at specific angles, as illustrated in [\[link\]](#). When the moon reflects from a lake, as shown in [\[link\]](#), a combination of these effects takes place.

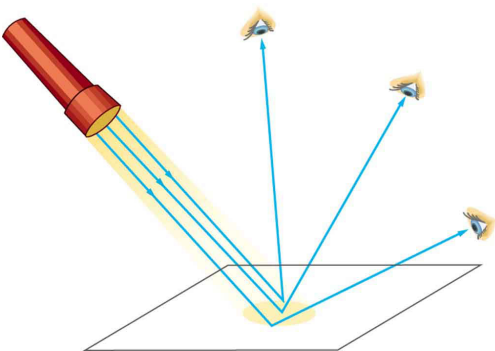


The law of reflection states that the angle of reflection equals the angle of incidence— $\theta_r = \theta_i$. The angles are measured relative to the perpendicular to

the surface at the point
where the ray strikes
the surface.

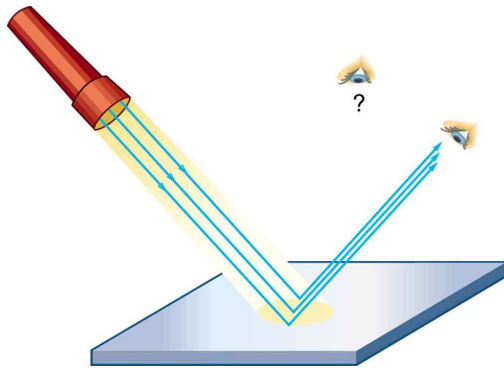


Light is diffused when it
reflects from a rough
surface. Here many
parallel rays are incident,
but they are reflected at
many different angles
since the surface is rough.

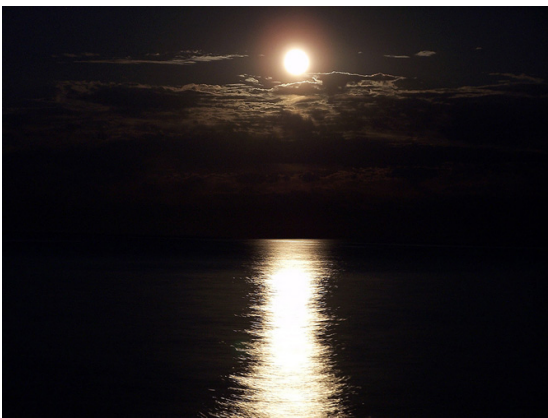


When a sheet of paper is
illuminated with many
parallel incident rays, it
can be seen at many
different angles, because

its surface is rough and
diffuses the light.



A mirror illuminated by
many parallel rays
reflects them in only one
direction, since its surface
is very smooth. Only the
observer at a particular
angle will see the
reflected light.



Moonlight is spread out
when it is reflected by the
lake, since the surface is
shiny but uneven. (credit:

Diego Torres Silvestre,
Flickr)

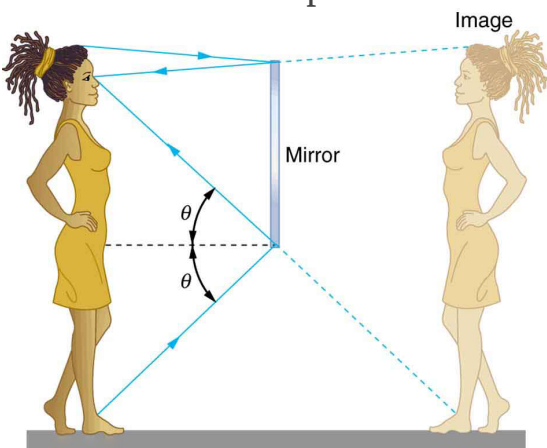
The law of reflection is very simple: The angle of reflection equals the angle of incidence.

Note:

The Law of Reflection

The angle of reflection equals the angle of incidence.

When we see ourselves in a mirror, it appears that our image is actually behind the mirror. This is illustrated in [\[link\]](#). We see the light coming from a direction determined by the law of reflection. The angles are such that our image is exactly the same distance behind the mirror as we stand away from the mirror. If the mirror is on the wall of a room, the images in it are all behind the mirror, which can make the room seem bigger. Although these mirror images make objects appear to be where they cannot be (like behind a solid wall), the images are not figments of our imagination. Mirror images can be photographed and videotaped by instruments and look just as they do with our eyes (optical instruments themselves). The precise manner in which images are formed by mirrors and lenses will be treated in later sections of this chapter.



Our image in a mirror is behind the mirror. The two rays shown are those that strike the mirror at just the correct angles to be reflected into the eyes of the person. The image appears to be in the direction the rays are coming from when they enter the eyes.

Note:**Take-Home Experiment: Law of Reflection**

Take a piece of paper and shine a flashlight at an angle at the paper, as shown in [\[link\]](#). Now shine the flashlight at a mirror at an angle. Do your observations confirm the predictions in [\[link\]](#) and [\[link\]](#)? Shine the flashlight on various surfaces and determine whether the reflected light is diffuse or not. You can choose a shiny metallic lid of a pot or your skin. Using the mirror and flashlight, can you confirm the law of reflection? You will need to draw lines on a piece of paper showing the incident and reflected rays. (This part works even better if you use a laser pencil.)

Section Summary

- The angle of reflection equals the angle of incidence.
- A mirror has a smooth surface and reflects light at specific angles.
- Light is diffused when it reflects from a rough surface.
- Mirror images can be photographed and videotaped by instruments.

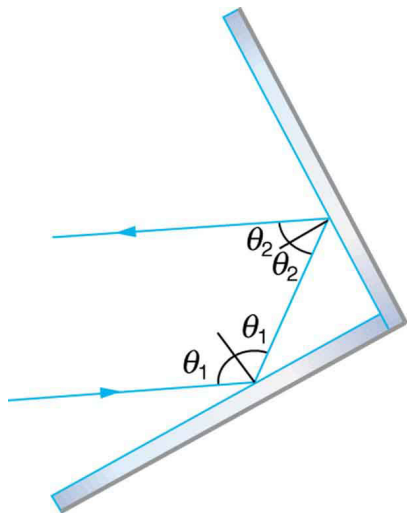
Conceptual Questions

Exercise:**Problem:**

Using the law of reflection, explain how powder takes the shine off of a person's nose. What is the name of the optical effect?

Problems & Exercises**Exercise:****Problem:**

Show that when light reflects from two mirrors that meet each other at a right angle, the outgoing ray is parallel to the incoming ray, as illustrated in the following figure.



A corner reflector sends the reflected ray back in a direction parallel to the incident ray, independent of incoming direction.

Exercise:

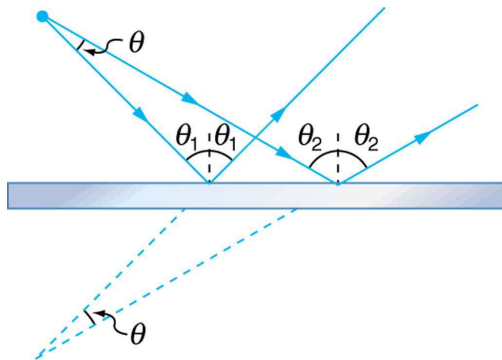
Problem:

Light shows staged with lasers use moving mirrors to swing beams and create colorful effects. Show that a light ray reflected from a mirror changes direction by 2θ when the mirror is rotated by an angle θ .

Exercise:

Problem:

A flat mirror is neither converging nor diverging. To prove this, consider two rays originating from the same point and diverging at an angle θ . Show that after striking a plane mirror, the angle between their directions remains θ .



A flat mirror neither converges nor diverges light rays. Two rays continue to diverge at the same angle after reflection.

Glossary

mirror

smooth surface that reflects light at specific angles, forming an image of the person or object in front of it

law of reflection

angle of reflection equals the angle of incidence

The Law of Refraction

- Determine the index of refraction, given the speed of light in a medium.

It is easy to notice some odd things when looking into a fish tank. For example, you may see the same fish appearing to be in two different places. (See [\[link\]](#).) This is because light coming from the fish to us changes direction when it leaves the tank, and in this case, it can travel two different paths to get to our eyes. The changing of a light ray's direction (loosely called bending) when it passes through variations in matter is called **refraction**. Refraction is responsible for a tremendous range of optical phenomena, from the action of lenses to voice transmission through optical fibers.

Note:

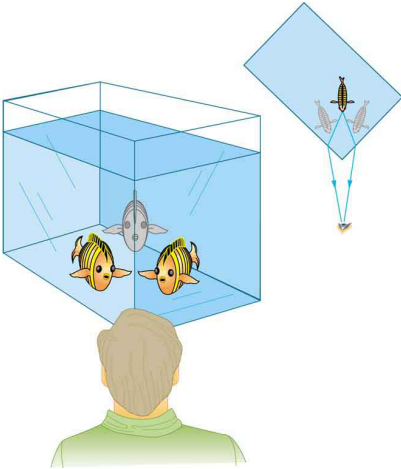
Refraction

The changing of a light ray's direction (loosely called bending) when it passes through variations in matter is called refraction.

Note:

Speed of Light

The speed of light c not only affects refraction, it is one of the central concepts of Einstein's theory of relativity. As the accuracy of the measurements of the speed of light were improved, c was found not to depend on the velocity of the source or the observer. However, the speed of light does vary in a precise manner with the material it traverses. These facts have far-reaching implications, as we will see in [Special Relativity](#). It makes connections between space and time and alters our expectations that all observers measure the same time for the same event, for example. The speed of light is so important that its value in a vacuum is one of the most fundamental constants in nature as well as being one of the four fundamental SI units.



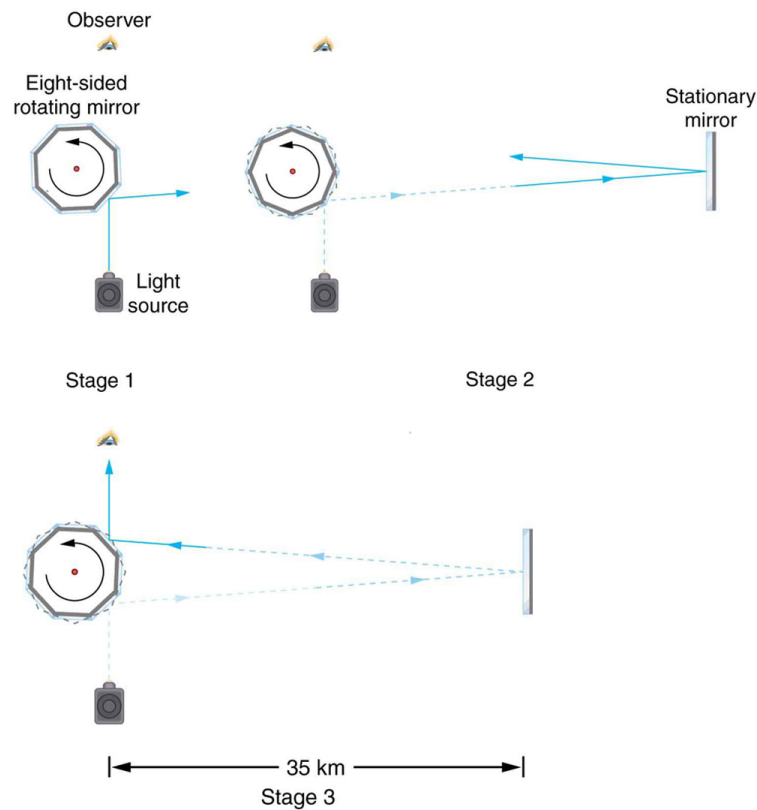
Looking at the fish tank as shown, we can see the same fish in two different locations, because light changes directions when it passes from water to air. In this case, the light can reach the observer by two different paths, and so the fish seems to be in two different places. This bending of light is called refraction and is responsible for many optical phenomena.

Why does light change direction when passing from one material (medium) to another? It is because light changes speed when going from one material

to another. So before we study the law of refraction, it is useful to discuss the speed of light and how it varies in different media.

The Speed of Light

Early attempts to measure the speed of light, such as those made by Galileo, determined that light moved extremely fast, perhaps instantaneously. The first real evidence that light traveled at a finite speed came from the Danish astronomer Ole Roemer in the late 17th century. Roemer had noted that the average orbital period of one of Jupiter's moons, as measured from Earth, varied depending on whether Earth was moving toward or away from Jupiter. He correctly concluded that the apparent change in period was due to the change in distance between Earth and Jupiter and the time it took light to travel this distance. From his 1676 data, a value of the speed of light was calculated to be 2.26×10^8 m/s (only 25% different from today's accepted value). In more recent times, physicists have measured the speed of light in numerous ways and with increasing accuracy. One particularly direct method, used in 1887 by the American physicist Albert Michelson (1852–1931), is illustrated in [\[link\]](#). Light reflected from a rotating set of mirrors was reflected from a stationary mirror 35 km away and returned to the rotating mirrors. The time for the light to travel can be determined by how fast the mirrors must rotate for the light to be returned to the observer's eye.



A schematic of early apparatus used by Michelson and others to determine the speed of light. As the mirrors rotate, the reflected ray is only briefly directed at the stationary mirror. The returning ray will be reflected into the observer's eye only if the next mirror has rotated into the correct position just as the ray returns. By measuring the correct rotation rate, the time for the round trip can be measured and the speed of light calculated. Michelson's calculated value of the speed of light was only 0.04% different from the value used today.

The speed of light is now known to great precision. In fact, the speed of light in a vacuum c is so important that it is accepted as one of the basic physical quantities and has the fixed value

Equation:

$$c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s},$$

where the approximate value of $3.00 \times 10^8 \text{ m/s}$ is used whenever three-digit accuracy is sufficient. The speed of light through matter is less than it is in a vacuum, because light interacts with atoms in a material. The speed of light depends strongly on the type of material, since its interaction with different atoms, crystal lattices, and other substructures varies. We define the **index of refraction** n of a material to be

Equation:

$$n = \frac{c}{v},$$

where v is the observed speed of light in the material. Since the speed of light is always less than c in matter and equals c only in a vacuum, the index of refraction is always greater than or equal to one.

Note:

Value of the Speed of Light

Equation:

$$c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}$$

Note:

Index of Refraction

Equation:

$$n = \frac{c}{v}$$

That is, $n \geq 1$. [\[link\]](#) gives the indices of refraction for some representative substances. The values are listed for a particular wavelength of light, because they vary slightly with wavelength. (This can have important effects, such as colors produced by a prism.) Note that for gases, n is close to 1.0. This seems reasonable, since atoms in gases are widely separated and light travels at c in the vacuum between atoms. It is common to take $n = 1$ for gases unless great precision is needed. Although the speed of light v in a medium varies considerably from its value c in a vacuum, it is still a large speed.

Medium	n
<i>Gases at 0°C, 1 atm</i>	
Air	1.000293
Carbon dioxide	1.00045
Hydrogen	1.000139
Oxygen	1.000271
<i>Liquids at 20°C</i>	
Benzene	1.501
Carbon disulfide	1.628

Medium	<i>n</i>
Carbon tetrachloride	1.461
Ethanol	1.361
Glycerine	1.473
Water, fresh	1.333
<i>Solids at 20°C</i>	
Diamond	2.419
Fluorite	1.434
Glass, crown	1.52
Glass, flint	1.66
Ice at 20°C	1.309
Polystyrene	1.49
Plexiglas	1.51
Quartz, crystalline	1.544
Quartz, fused	1.458
Sodium chloride	1.544
Zircon	1.923

Index of Refraction in Various Media

Example:**Speed of Light in Matter**

Calculate the speed of light in zircon, a material used in jewelry to imitate diamond.

Strategy

The speed of light in a material, v , can be calculated from the index of refraction n of the material using the equation $n = c/v$.

Solution

The equation for index of refraction states that $n = c/v$. Rearranging this to determine v gives

Equation:

$$v = \frac{c}{n}.$$

The index of refraction for zircon is given as 1.923 in [\[link\]](#), and c is given in the equation for speed of light. Entering these values in the last expression gives

Equation:

$$\begin{aligned} v &= \frac{3.00 \times 10^8 \text{ m/s}}{1.923} \\ &= 1.56 \times 10^8 \text{ m/s.} \end{aligned}$$

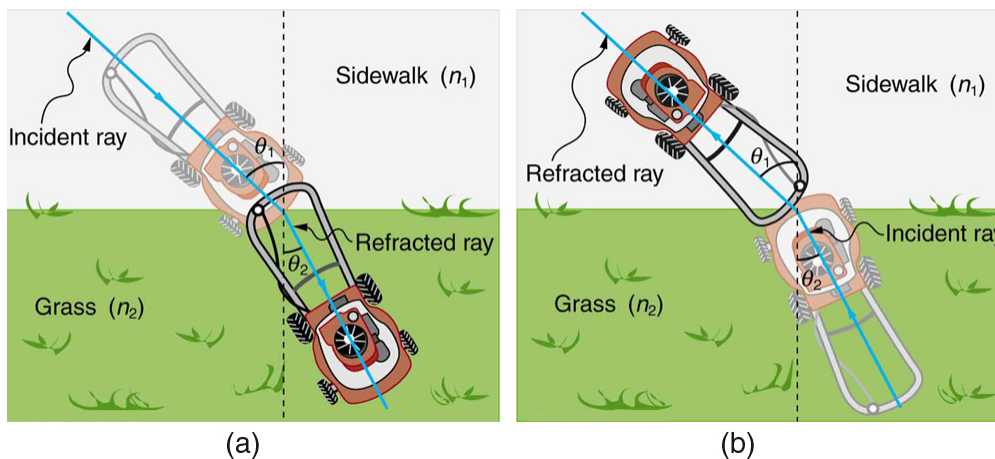
Discussion

This speed is slightly larger than half the speed of light in a vacuum and is still high compared with speeds we normally experience. The only substance listed in [\[link\]](#) that has a greater index of refraction than zircon is diamond. We shall see later that the large index of refraction for zircon makes it sparkle more than glass, but less than diamond.

Law of Refraction

[\[link\]](#) shows how a ray of light changes direction when it passes from one medium to another. As before, the angles are measured relative to a perpendicular to the surface at the point where the light ray crosses it.

(Some of the incident light will be reflected from the surface, but for now we will concentrate on the light that is transmitted.) The change in direction of the light ray depends on how the speed of light changes. The change in the speed of light is related to the indices of refraction of the media involved. In the situations shown in [\[link\]](#), medium 2 has a greater index of refraction than medium 1. This means that the speed of light is less in medium 2 than in medium 1. Note that as shown in [\[link\]](#)(a), the direction of the ray moves closer to the perpendicular when it slows down. Conversely, as shown in [\[link\]](#)(b), the direction of the ray moves away from the perpendicular when it speeds up. The path is exactly reversible. In both cases, you can imagine what happens by thinking about pushing a lawn mower from a footpath onto grass, and vice versa. Going from the footpath to grass, the front wheels are slowed and pulled to the side as shown. This is the same change in direction as for light when it goes from a fast medium to a slow one. When going from the grass to the footpath, the front wheels can move faster and the mower changes direction as shown. This, too, is the same change in direction as for light going from slow to fast.



The change in direction of a light ray depends on how the speed of light changes when it crosses from one medium to another. The speed of light is greater in medium 1 than in medium 2 in the situations shown here. (a) A ray of light moves closer to the perpendicular when it slows down. This is analogous to what happens when a lawn mower goes from a footpath to grass. (b) A ray of

light moves away from the perpendicular when it speeds up. This is analogous to what happens when a lawn mower goes from grass to footpath. The paths are exactly reversible.

The amount that a light ray changes its direction depends both on the incident angle and the amount that the speed changes. For a ray at a given incident angle, a large change in speed causes a large change in direction, and thus a large change in angle. The exact mathematical relationship is the **law of refraction**, or “Snell’s Law,” which is stated in equation form as **Equation:**

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

Here n_1 and n_2 are the indices of refraction for medium 1 and 2, and θ_1 and θ_2 are the angles between the rays and the perpendicular in medium 1 and 2, as shown in [\[link\]](#). The incoming ray is called the incident ray and the outgoing ray the refracted ray, and the associated angles the incident angle and the refracted angle. The law of refraction is also called Snell’s law after the Dutch mathematician Willebrord Snell (1591–1626), who discovered it in 1621. Snell’s experiments showed that the law of refraction was obeyed and that a characteristic index of refraction n could be assigned to a given medium. Snell was not aware that the speed of light varied in different media, but through experiments he was able to determine indices of refraction from the way light rays changed direction.

Note:

The Law of Refraction

Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

Note:**Take-Home Experiment: A Broken Pencil**

A classic observation of refraction occurs when a pencil is placed in a glass half filled with water. Do this and observe the shape of the pencil when you look at the pencil sideways, that is, through air, glass, water. Explain your observations. Draw ray diagrams for the situation.

Example:**Determine the Index of Refraction from Refraction Data**

Find the index of refraction for medium 2 in [\[link\]](#)(a), assuming medium 1 is air and given the incident angle is 30.0° and the angle of refraction is 22.0° .

Strategy

The index of refraction for air is taken to be 1 in most cases (and up to four significant figures, it is 1.000). Thus $n_1 = 1.00$ here. From the given information, $\theta_1 = 30.0^\circ$ and $\theta_2 = 22.0^\circ$. With this information, the only unknown in Snell's law is n_2 , so that it can be used to find this unknown.

Solution

Snell's law is

Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

Rearranging to isolate n_2 gives

Equation:

$$n_2 = n_1 \frac{\sin \theta_1}{\sin \theta_2}.$$

Entering known values,

Equation:

$$\begin{aligned} n_2 &= 1.00 \frac{\sin 30.0^\circ}{\sin 22.0^\circ} = \frac{0.500}{0.375} \\ &= 1.33. \end{aligned}$$

Discussion

This is the index of refraction for water, and Snell could have determined it by measuring the angles and performing this calculation. He would then have found 1.33 to be the appropriate index of refraction for water in all other situations, such as when a ray passes from water to glass. Today we can verify that the index of refraction is related to the speed of light in a medium by measuring that speed directly.

Example:**A Larger Change in Direction**

Suppose that in a situation like that in [\[link\]](#), light goes from air to diamond and that the incident angle is 30.0° . Calculate the angle of refraction θ_2 in the diamond.

Strategy

Again the index of refraction for air is taken to be $n_1 = 1.00$, and we are given $\theta_1 = 30.0^\circ$. We can look up the index of refraction for diamond in [\[link\]](#), finding $n_2 = 2.419$. The only unknown in Snell's law is θ_2 , which we wish to determine.

Solution

Solving Snell's law for $\sin \theta_2$ yields

Equation:

$$\sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1.$$

Entering known values,

Equation:

$$\sin \theta_2 = \frac{1.00}{2.419} \sin 30.0^\circ = (0.413)(0.500) = 0.207.$$

The angle is thus

Equation:

$$\theta_2 = \sin^{-1} 0.207 = 11.9^\circ.$$

Discussion

For the same 30° angle of incidence, the angle of refraction in diamond is significantly smaller than in water (11.9° rather than 22° —see the preceding example). This means there is a larger change in direction in diamond. The cause of a large change in direction is a large change in the index of refraction (or speed). In general, the larger the change in speed, the greater the effect on the direction of the ray.

Section Summary

- The changing of a light ray's direction when it passes through variations in matter is called refraction.
- The speed of light in vacuum
 $c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}$.
- Index of refraction $n = \frac{c}{v}$, where v is the speed of light in the material, c is the speed of light in vacuum, and n is the index of refraction.
- Snell's law, the law of refraction, is stated in equation form as $n_1 \sin \theta_1 = n_2 \sin \theta_2$.

Conceptual Questions

Exercise:

Problem:

Diffusion by reflection from a rough surface is described in this chapter. Light can also be diffused by refraction. Describe how this occurs in a specific situation, such as light interacting with crushed ice.

Exercise:

Problem:

Why is the index of refraction always greater than or equal to 1?

Exercise:

Problem:

Does the fact that the light flash from lightning reaches you before its sound prove that the speed of light is extremely large or simply that it is greater than the speed of sound? Discuss how you could use this effect to get an estimate of the speed of light.

Exercise:**Problem:**

Will light change direction toward or away from the perpendicular when it goes from air to water? Water to glass? Glass to air?

Exercise:**Problem:**

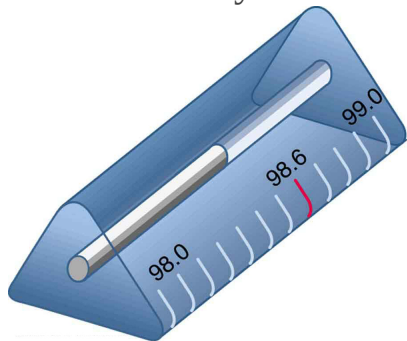
Explain why an object in water always appears to be at a depth shallower than it actually is? Why do people sometimes sustain neck and spinal injuries when diving into unfamiliar ponds or waters?

Exercise:**Problem:**

Explain why a person's legs appear very short when wading in a pool. Justify your explanation with a ray diagram showing the path of rays from the feet to the eye of an observer who is out of the water.

Exercise:

Problem: Why is the front surface of a thermometer curved as shown?



The curved surface
of the thermometer
serves a purpose.

Exercise:

Problem:

Suppose light were incident from air onto a material that had a negative index of refraction, say -1.3 ; where does the refracted light ray go?

Problems & Exercises

Exercise:

Problem: What is the speed of light in water? In glycerine?

Solution:

2.25×10^8 m/s in water

2.04×10^8 m/s in glycerine

Exercise:

Problem: What is the speed of light in air? In crown glass?

Exercise:

Problem:

Calculate the index of refraction for a medium in which the speed of light is 2.012×10^8 m/s, and identify the most likely substance based on [\[link\]](#).

Solution:

1.490, polystyrene

Exercise:**Problem:**

In what substance in [\[link\]](#) is the speed of light 2.290×10^8 m/s?

Exercise:**Problem:**

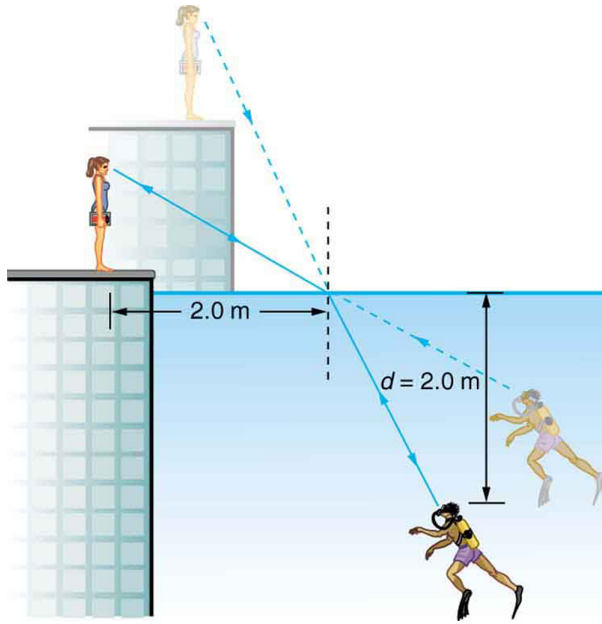
There was a major collision of an asteroid with the Moon in medieval times. It was described by monks at Canterbury Cathedral in England as a red glow on and around the Moon. How long after the asteroid hit the Moon, which is 3.84×10^5 km away, would the light first arrive on Earth?

Solution:

1.28 s

Exercise:**Problem:**

A scuba diver training in a pool looks at his instructor as shown in [\[link\]](#). What angle does the ray from the instructor's face make with the perpendicular to the water at the point where the ray enters? The angle between the ray in the water and the perpendicular to the water is 25.0° .



A scuba diver in a pool and his trainer look at each other.

Exercise:

Problem:

Components of some computers communicate with each other through optical fibers having an index of refraction $n = 1.55$. What time in nanoseconds is required for a signal to travel 0.200 m through such a fiber?

Solution:

1.03 ns

Exercise:

Problem:

(a) Given that the angle between the ray in the water and the perpendicular to the water is 25.0° , and using information in [\[link\]](#), find the height of the instructor's head above the water, noting that you will first have to calculate the angle of incidence. (b) Find the apparent depth of the diver's head below water as seen by the instructor.

Exercise:**Problem:**

Suppose you have an unknown clear substance immersed in water, and you wish to identify it by finding its index of refraction. You arrange to have a beam of light enter it at an angle of 45.0° , and you observe the angle of refraction to be 40.3° . What is the index of refraction of the substance and its likely identity?

Solution:

$n = 1.46$, fused quartz

Exercise:**Problem:**

On the Moon's surface, lunar astronauts placed a corner reflector, off which a laser beam is periodically reflected. The distance to the Moon is calculated from the round-trip time. What percent correction is needed to account for the delay in time due to the slowing of light in Earth's atmosphere? Assume the distance to the Moon is precisely 3.84×10^8 m, and Earth's atmosphere (which varies in density with altitude) is equivalent to a layer 30.0 km thick with a constant index of refraction $n = 1.000293$.

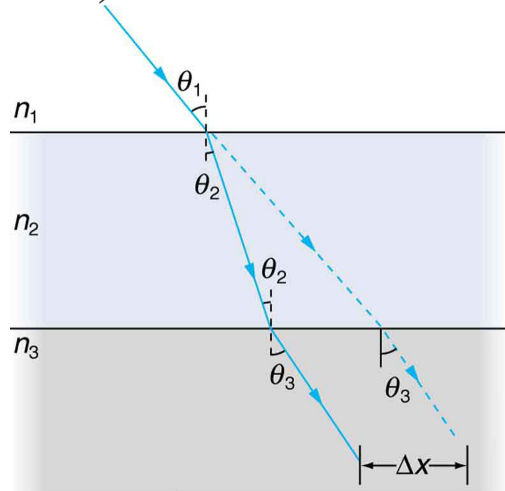
Exercise:

Problem:

Suppose [\[link\]](#) represents a ray of light going from air through crown glass into water, such as going into a fish tank. Calculate the amount the ray is displaced by the glass (Δx), given that the incident angle is 40.0° and the glass is 1.00 cm thick.

Exercise:**Problem:**

[\[link\]](#) shows a ray of light passing from one medium into a second and then a third. Show that θ_3 is the same as it would be if the second medium were not present (provided total internal reflection does not occur).



A ray of light passes from one medium to a third by traveling through a second. The final direction is the same as if the second medium were not present, but the ray is displaced by Δx (shown exaggerated).

Exercise:**Problem: Unreasonable Results**

Suppose light travels from water to another substance, with an angle of incidence of 10.0° and an angle of refraction of 14.9° . (a) What is the index of refraction of the other substance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) 0.898

(b) Can't have $n < 1.00$ since this would imply a speed greater than c .

(c) Refracted angle is too big relative to the angle of incidence.

Exercise:**Problem: Construct Your Own Problem**

Consider sunlight entering the Earth's atmosphere at sunrise and sunset—that is, at a 90° incident angle. Taking the boundary between nearly empty space and the atmosphere to be sudden, calculate the angle of refraction for sunlight. This lengthens the time the Sun appears to be above the horizon, both at sunrise and sunset. Now construct a problem in which you determine the angle of refraction for different models of the atmosphere, such as various layers of varying density. Your instructor may wish to guide you on the level of complexity to consider and on how the index of refraction varies with air density.

Exercise:**Problem: Unreasonable Results**

Light traveling from water to a gemstone strikes the surface at an angle of 80.0° and has an angle of refraction of 15.2° . (a) What is the speed

of light in the gemstone? (b) What is unreasonable about this result?
(c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $\frac{c}{5.00}$

(b) Speed of light too slow, since index is much greater than that of diamond.

(c) Angle of refraction is unreasonable relative to the angle of incidence.

Glossary

refraction

changing of a light ray's direction when it passes through variations in matter

index of refraction

for a material, the ratio of the speed of light in vacuum to that in the material

Total Internal Reflection

- Explain the phenomenon of total internal reflection.
- Describe the workings and uses of fiber optics.
- Analyze the reason for the sparkle of diamonds.

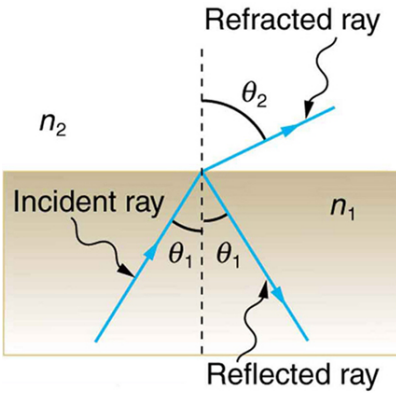
A good-quality mirror may reflect more than 90% of the light that falls on it, absorbing the rest. But it would be useful to have a mirror that reflects all of the light that falls on it. Interestingly, we can produce *total reflection* using an aspect of *refraction*.

Consider what happens when a ray of light strikes the surface between two materials, such as is shown in [\[link\]](#)(a). Part of the light crosses the boundary and is refracted; the rest is reflected. If, as shown in the figure, the index of refraction for the second medium is less than for the first, the ray bends away from the perpendicular. (Since $n_1 > n_2$, the angle of refraction is greater than the angle of incidence—that is, $\theta_2 > \theta_1$.) Now imagine what happens as the incident angle is increased. This causes θ_2 to increase also. The largest the angle of refraction θ_2 can be is 90° , as shown in [\[link\]](#)(b). The **critical angle** θ_c for a combination of materials is defined to be the incident angle θ_1 that produces an angle of refraction of 90° . That is, θ_c is the incident angle for which $\theta_2 = 90^\circ$. If the incident angle θ_1 is greater than the critical angle, as shown in [\[link\]](#)(c), then all of the light is reflected back into medium 1, a condition called **total internal reflection**.

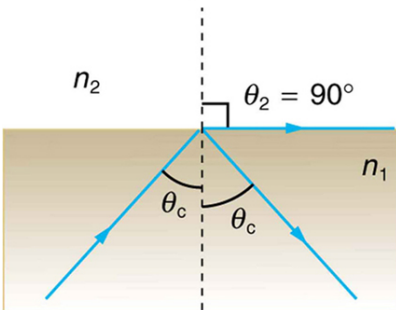
Note:

Critical Angle

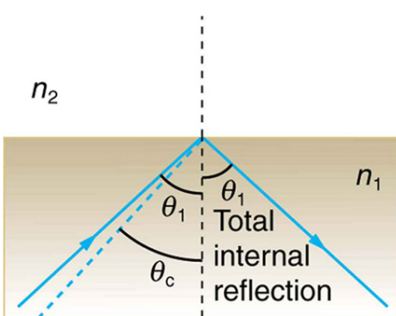
The incident angle θ_1 that produces an angle of refraction of 90° is called the critical angle, θ_c .



(a)



(b)



(c)

(a) A ray of light crosses a boundary where the speed of light increases and the index of refraction decreases. That is, $n_2 < n_1$. The ray bends away from the perpendicular.

(b) The critical

angle θ_c is the one for which the angle of refraction is 90° . (c)

Total internal reflection occurs when the incident angle is greater than the critical angle.

Snell's law states the relationship between angles and indices of refraction. It is given by

Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

When the incident angle equals the critical angle ($\theta_1 = \theta_c$), the angle of refraction is 90° ($\theta_2 = 90^\circ$). Noting that $\sin 90^\circ = 1$, Snell's law in this case becomes

Equation:

$$n_1 \sin \theta_1 = n_2.$$

The critical angle θ_c for a given combination of materials is thus

Equation:

$$\theta_c = \sin^{-1}(n_2/n_1) \text{ for } n_1 > n_2.$$

Total internal reflection occurs for any incident angle greater than the critical angle θ_c , and it can only occur when the second medium has an index of refraction less than the first. Note the above equation is written for a light ray that travels in medium 1 and reflects from medium 2, as shown in the figure.

Example:**How Big is the Critical Angle Here?**

What is the critical angle for light traveling in a polystyrene (a type of plastic) pipe surrounded by air?

Strategy

The index of refraction for polystyrene is found to be 1.49 in [\[link\]](#), and the index of refraction of air can be taken to be 1.00, as before. Thus, the condition that the second medium (air) has an index of refraction less than the first (plastic) is satisfied, and the equation $\theta_c = \sin^{-1}(n_2/n_1)$ can be used to find the critical angle θ_c . Here, then, $n_2 = 1.00$ and $n_1 = 1.49$.

Solution

The critical angle is given by

Equation:

$$\theta_c = \sin^{-1}(n_2/n_1).$$

Substituting the identified values gives

Equation:

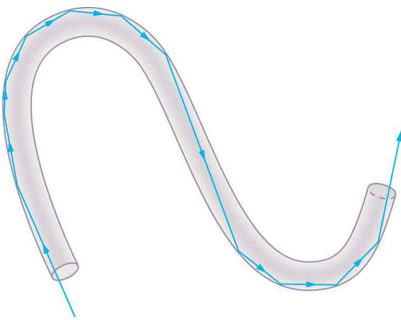
$$\theta_c = \sin^{-1}(1.00/1.49) = \sin^{-1}(0.671) \\ 42.2^\circ.$$

Discussion

This means that any ray of light inside the plastic that strikes the surface at an angle greater than 42.2° will be totally reflected. This will make the inside surface of the clear plastic a perfect mirror for such rays without any need for the silvering used on common mirrors. Different combinations of materials have different critical angles, but any combination with $n_1 > n_2$ can produce total internal reflection. The same calculation as made here shows that the critical angle for a ray going from water to air is 48.6° , while that from diamond to air is 24.4° , and that from flint glass to crown glass is 66.3° . There is no total reflection for rays going in the other direction—for example, from air to water—since the condition that the second medium must have a smaller index of refraction is not satisfied. A number of interesting applications of total internal reflection follow.

Fiber Optics: Endoscopes to Telephones

Fiber optics is one application of total internal reflection that is in wide use. In communications, it is used to transmit telephone, internet, and cable TV signals. **Fiber optics** employs the transmission of light down fibers of plastic or glass. Because the fibers are thin, light entering one is likely to strike the inside surface at an angle greater than the critical angle and, thus, be totally reflected (See [\[link\]](#).) The index of refraction outside the fiber must be smaller than inside, a condition that is easily satisfied by coating the outside of the fiber with a material having an appropriate refractive index. In fact, most fibers have a varying refractive index to allow more light to be guided along the fiber through total internal reflection. Rays are reflected around corners as shown, making the fibers into tiny light pipes.

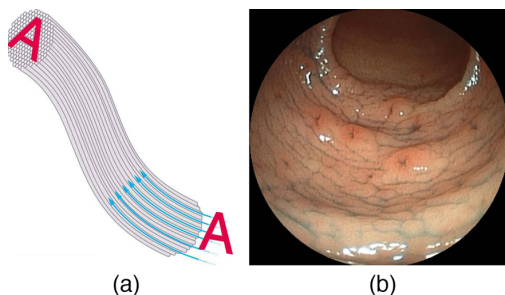


Light entering a thin fiber may strike the inside surface at large or grazing angles and is completely reflected if these angles exceed the critical angle. Such rays continue down the fiber, even following it around corners, since the angles of reflection

and incidence
remain large.

Bundles of fibers can be used to transmit an image without a lens, as illustrated in [\[link\]](#). The output of a device called an **endoscope** is shown in [\[link\]](#)(b). Endoscopes are used to explore the body through various orifices or minor incisions. Light is transmitted down one fiber bundle to illuminate internal parts, and the reflected light is transmitted back out through another to be observed. Surgery can be performed, such as arthroscopic surgery on the knee joint, employing cutting tools attached to and observed with the endoscope. Samples can also be obtained, such as by lassoing an intestinal polyp for external examination.

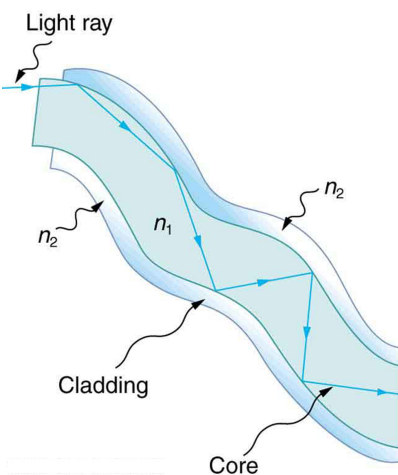
Fiber optics has revolutionized surgical techniques and observations within the body. There are a host of medical diagnostic and therapeutic uses. The flexibility of the fiber optic bundle allows it to navigate around difficult and small regions in the body, such as the intestines, the heart, blood vessels, and joints. Transmission of an intense laser beam to burn away obstructing plaques in major arteries as well as delivering light to activate chemotherapy drugs are becoming commonplace. Optical fibers have in fact enabled microsurgery and remote surgery where the incisions are small and the surgeon's fingers do not need to touch the diseased tissue.



(a) An image is
transmitted by a bundle of
fibers that have fixed

neighbors. (b) An endoscope is used to probe the body, both transmitting light to the interior and returning an image such as the one shown. (credit: Med_Chaos, Wikimedia Commons)

Fibers in bundles are surrounded by a cladding material that has a lower index of refraction than the core. (See [\[link\]](#).) The cladding prevents light from being transmitted between fibers in a bundle. Without cladding, light could pass between fibers in contact, since their indices of refraction are identical. Since no light gets into the cladding (there is total internal reflection back into the core), none can be transmitted between clad fibers that are in contact with one another. The cladding prevents light from escaping out of the fiber; instead most of the light is propagated along the length of the fiber, minimizing the loss of signal and ensuring that a quality image is formed at the other end. The cladding and an additional protective layer make optical fibers flexible and durable.



Fibers in bundles
are clad by a
material that has a
lower index of
refraction than the
core to ensure total
internal reflection,
even when fibers
are in contact with
one another. This
shows a single fiber
with its cladding.

Note:

Cladding

The cladding prevents light from being transmitted between fibers in a bundle.

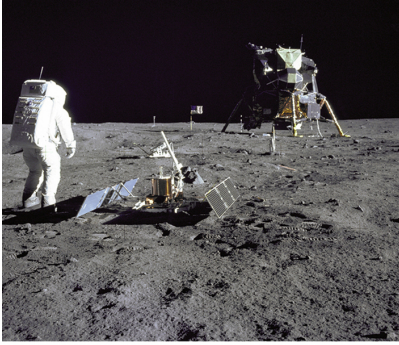
Special tiny lenses that can be attached to the ends of bundles of fibers are being designed and fabricated. Light emerging from a fiber bundle can be focused and a tiny spot can be imaged. In some cases the spot can be scanned, allowing quality imaging of a region inside the body. Special minute optical filters inserted at the end of the fiber bundle have the capacity to image tens of microns below the surface without cutting the surface—non-intrusive diagnostics. This is particularly useful for determining the extent of cancers in the stomach and bowel.

Most telephone conversations and Internet communications are now carried by laser signals along optical fibers. Extensive optical fiber cables have been placed on the ocean floor and underground to enable optical communications. Optical fiber communication systems offer several advantages over electrical (copper) based systems, particularly for long

distances. The fibers can be made so transparent that light can travel many kilometers before it becomes dim enough to require amplification—much superior to copper conductors. This property of optical fibers is called *low loss*. Lasers emit light with characteristics that allow far more conversations in one fiber than are possible with electric signals on a single conductor. This property of optical fibers is called *high bandwidth*. Optical signals in one fiber do not produce undesirable effects in other adjacent fibers. This property of optical fibers is called *reduced crosstalk*. We shall explore the unique characteristics of laser radiation in a later chapter.

Corner Reflectors and Diamonds

A light ray that strikes an object consisting of two mutually perpendicular reflecting surfaces is reflected back exactly parallel to the direction from which it came. This is true whenever the reflecting surfaces are perpendicular, and it is independent of the angle of incidence. Such an object, shown in [\[link\]](#), is called a **corner reflector**, since the light bounces from its inside corner. Many inexpensive reflector buttons on bicycles, cars, and warning signs have corner reflectors designed to return light in the direction from which it originated. It was more expensive for astronauts to place one on the moon. Laser signals can be bounced from that corner reflector to measure the gradually increasing distance to the moon with great precision.



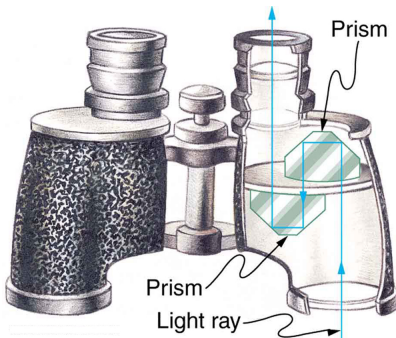
(a)



(b)

(a) Astronauts placed a corner reflector on the moon to measure its gradually increasing orbital distance. (credit: NASA) (b) The bright spots on these bicycle safety reflectors are reflections of the flash of the camera that took this picture on a dark night. (credit: Julo, Wikimedia Commons)

Corner reflectors are perfectly efficient when the conditions for total internal reflection are satisfied. With common materials, it is easy to obtain a critical angle that is less than 45° . One use of these perfect mirrors is in binoculars, as shown in [\[link\]](#). Another use is in periscopes found in submarines.

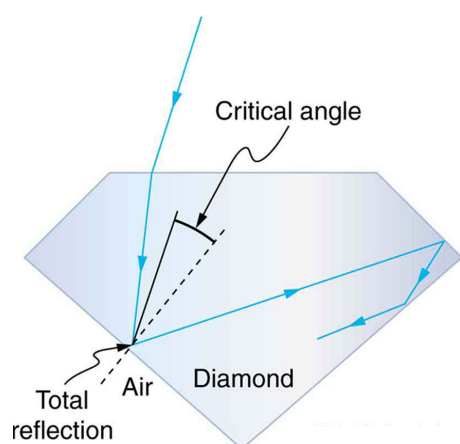


These binoculars employ corner reflectors with total internal reflection to get light to the observer's eyes.

The Sparkle of Diamonds

Total internal reflection, coupled with a large index of refraction, explains why diamonds sparkle more than other materials. The critical angle for a diamond-to-air surface is only 24.4° , and so when light enters a diamond, it has trouble getting back out. (See [\[link\]](#).) Although light freely enters the diamond, it can exit only if it makes an angle less than 24.4° . Facets on diamonds are specifically intended to make this unlikely, so that the light can exit only in certain places. Good diamonds are very clear, so that the light makes many internal reflections and is concentrated at the few places it can exit—hence the sparkle. (Zircon is a natural gemstone that has an exceptionally large index of refraction, but not as large as diamond, so it is

not as highly prized. Cubic zirconia is manufactured and has an even higher index of refraction (≈ 2.17), but still less than that of diamond.) The colors you see emerging from a sparkling diamond are not due to the diamond's color, which is usually nearly colorless. Those colors result from dispersion, the topic of [Dispersion: The Rainbow and Prisms](#). Colored diamonds get their color from structural defects of the crystal lattice and the inclusion of minute quantities of graphite and other materials. The Argyle Mine in Western Australia produces around 90% of the world's pink, red, champagne, and cognac diamonds, while around 50% of the world's clear diamonds come from central and southern Africa.



Light cannot easily escape a diamond, because its critical angle with air is so small. Most reflections are total, and the facets are placed so that light can exit only in particular ways—thus concentrating the light and making the diamond sparkle.

Note:**PhET Explorations: Bending Light**

Explore bending of light between two media with different indices of refraction. See how changing from air to water to glass changes the bending angle. Play with prisms of different shapes and make rainbows.

https://phet.colorado.edu/sims/html/bending-light/latest/bending-light_en.html

Section Summary

- The incident angle that produces an angle of refraction of 90° is called critical angle.
- Total internal reflection is a phenomenon that occurs at the boundary between two mediums, such that if the incident angle in the first medium is greater than the critical angle, then all the light is reflected back into that medium.
- Fiber optics involves the transmission of light down fibers of plastic or glass, applying the principle of total internal reflection.
- Endoscopes are used to explore the body through various orifices or minor incisions, based on the transmission of light through optical fibers.
- Cladding prevents light from being transmitted between fibers in a bundle.
- Diamonds sparkle due to total internal reflection coupled with a large index of refraction.

Conceptual Questions

Exercise:**Problem:**

A ring with a colorless gemstone is dropped into water. The gemstone becomes invisible when submerged. Can it be a diamond? Explain.

Exercise:**Problem:**

A high-quality diamond may be quite clear and colorless, transmitting all visible wavelengths with little absorption. Explain how it can sparkle with flashes of brilliant color when illuminated by white light.

Exercise:**Problem:**

Is it possible that total internal reflection plays a role in rainbows? Explain in terms of indices of refraction and angles, perhaps referring to [\[link\]](#). Some of us have seen the formation of a double rainbow. Is it physically possible to observe a triple rainbow?



Double rainbows are not a very common observance. (credit: InvictusOU812, Flickr)

Exercise:

Problem:

The most common type of mirage is an illusion that light from faraway objects is reflected by a pool of water that is not really there. Mirages are generally observed in deserts, when there is a hot layer of air near the ground. Given that the refractive index of air is lower for air at higher temperatures, explain how mirages can be formed.

Problems & Exercises**Exercise:****Problem:**

Verify that the critical angle for light going from water to air is 48.6° , as discussed at the end of [\[link\]](#), regarding the critical angle for light traveling in a polystyrene (a type of plastic) pipe surrounded by air.

Exercise:**Problem:**

(a) At the end of [\[link\]](#), it was stated that the critical angle for light going from diamond to air is 24.4° . Verify this. (b) What is the critical angle for light going from zircon to air?

Exercise:**Problem:**

An optical fiber uses flint glass clad with crown glass. What is the critical angle?

Solution:

66.3°

Exercise:

Problem:

At what minimum angle will you get total internal reflection of light traveling in water and reflected from ice?

Exercise:**Problem:**

Suppose you are using total internal reflection to make an efficient corner reflector. If there is air outside and the incident angle is 45.0° , what must be the minimum index of refraction of the material from which the reflector is made?

Solution:

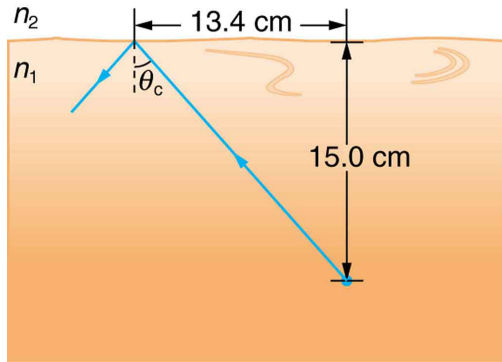
> 1.414

Exercise:**Problem:**

You can determine the index of refraction of a substance by determining its critical angle. (a) What is the index of refraction of a substance that has a critical angle of 68.4° when submerged in water? What is the substance, based on [\[link\]](#)? (b) What would the critical angle be for this substance in air?

Exercise:**Problem:**

A ray of light, emitted beneath the surface of an unknown liquid with air above it, undergoes total internal reflection as shown in [\[link\]](#). What is the index of refraction for the liquid and its likely identification?



A light ray inside a liquid strikes the surface at the critical angle and undergoes total internal reflection.

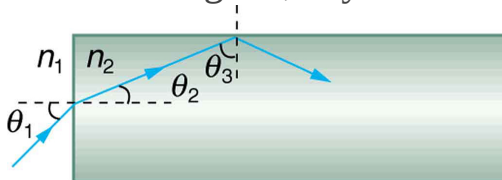
Solution:

1.50, benzene

Exercise:

Problem:

A light ray entering an optical fiber surrounded by air is first refracted and then reflected as shown in [\[link\]](#). Show that if the fiber is made from crown glass, any incident ray will be totally internally reflected.



A light ray enters the end of a fiber, the surface of which is perpendicular to its sides. Examine the conditions under which it

may be totally internally reflected.

Glossary

critical angle

incident angle that produces an angle of refraction of 90°

fiber optics

transmission of light down fibers of plastic or glass, applying the principle of total internal reflection

corner reflector

an object consisting of two mutually perpendicular reflecting surfaces, so that the light that enters is reflected back exactly parallel to the direction from which it came

zircon

natural gemstone with a large index of refraction

Dispersion: The Rainbow and Prisms

- Explain the phenomenon of dispersion and discuss its advantages and disadvantages.

Everyone enjoys the spectacle of a rainbow glimmering against a dark stormy sky. How does sunlight falling on clear drops of rain get broken into the rainbow of colors we see? The same process causes white light to be broken into colors by a clear glass prism or a diamond. (See [\[link\]](#).)



(a)



(b)

The colors of the rainbow (a) and those produced by a prism (b) are identical.

(credit: Alfredo55, Wikimedia Commons; NASA)

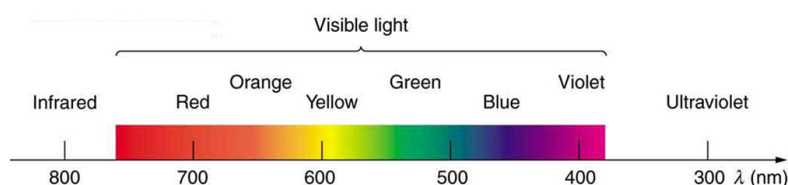
We see about six colors in a rainbow—red, orange, yellow, green, blue, and violet; sometimes indigo is listed, too. Those colors are associated with different wavelengths of light, as shown in [\[link\]](#). When our eye receives pure-wavelength light, we tend to see only one of the six colors, depending on wavelength. The thousands of other hues we can sense in other situations are our eye's response to various mixtures of wavelengths. White light, in particular, is a fairly uniform mixture of all visible wavelengths. Sunlight, considered to be white, actually appears to be a bit yellow because of its mixture of wavelengths, but it does contain all visible wavelengths. The sequence of colors in rainbows is the same sequence as the colors plotted versus wavelength in [\[link\]](#). What this implies is that white light is spread out according to

wavelength in a rainbow. **Dispersion** is defined as the spreading of white light into its full spectrum of wavelengths. More technically, dispersion occurs whenever there is a process that changes the direction of light in a manner that depends on wavelength. Dispersion, as a general phenomenon, can occur for any type of wave and always involves wavelength-dependent processes.

Note:

Dispersion

Dispersion is defined to be the spreading of white light into its full spectrum of wavelengths.



Even though rainbows are associated with seven colors, the rainbow is a continuous distribution of colors according to wavelengths.

Refraction is responsible for dispersion in rainbows and many other situations. The angle of refraction depends on the index of refraction, as we saw in [The Law of Refraction](#). We know that the index of refraction n depends on the medium. But for a given medium, n also depends on wavelength. (See [\[link\]](#). Note that, for a given medium, n increases as wavelength decreases and is greatest for violet light. Thus violet light is bent more than red light, as shown for a prism in [\[link\]\(b\)](#), and the light is dispersed into the same sequence of wavelengths as seen in [\[link\]](#) and [\[link\]](#).

Note:

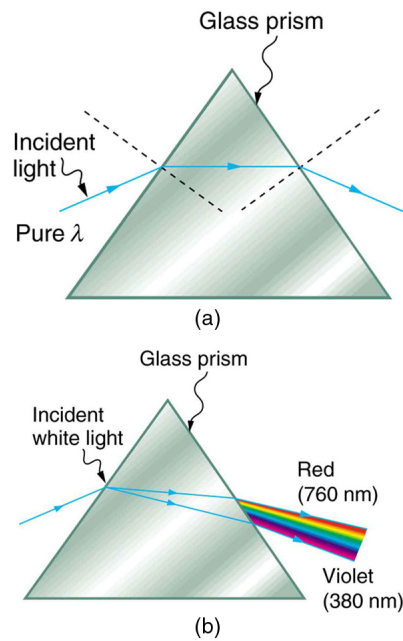
Making Connections: Dispersion

Any type of wave can exhibit dispersion. Sound waves, all types of electromagnetic waves, and water waves can be dispersed according to wavelength. Dispersion occurs whenever the speed of propagation depends on wavelength, thus separating and spreading out various wavelengths. Dispersion may require special circumstances and can result in spectacular displays such as in the production of a rainbow. This is also

true for sound, since all frequencies ordinarily travel at the same speed. If you listen to sound through a long tube, such as a vacuum cleaner hose, you can easily hear it is dispersed by interaction with the tube. Dispersion, in fact, can reveal a great deal about what the wave has encountered that disperses its wavelengths. The dispersion of electromagnetic radiation from outer space, for example, has revealed much about what exists between the stars—the so-called empty space.

Medium	Red (660 nm)	Orange (610 nm)	Yellow (580 nm)	Green (550 nm)	Blue (470 nm)	Violet (410 nm)
Water	1.331	1.332	1.333	1.335	1.338	1.342
Diamond	2.410	2.415	2.417	2.426	2.444	2.458
Glass, crown	1.512	1.514	1.518	1.519	1.524	1.530
Glass, flint	1.662	1.665	1.667	1.674	1.684	1.698
Polystyrene	1.488	1.490	1.492	1.493	1.499	1.506
Quartz, fused	1.455	1.456	1.458	1.459	1.462	1.468

Index of Refraction n in Selected Media at Various Wavelengths



(a) A pure wavelength of light falls onto a prism and is refracted at both surfaces. (b)

White light is dispersed by the prism (shown exaggerated).

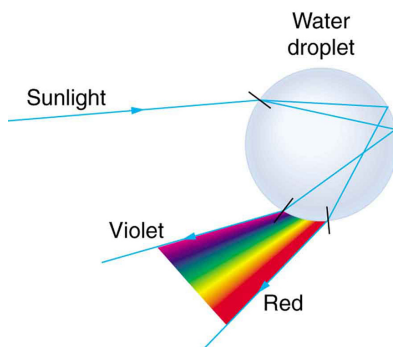
Since the index of refraction varies with wavelength, the angles of refraction vary with wavelength. A sequence of red to violet is produced, because the index of refraction increases steadily with decreasing wavelength.

Rainbows are produced by a combination of refraction and reflection. You may have noticed that you see a rainbow only when you look away from the sun. Light enters a drop of water and is reflected from the back of the drop, as shown in [\[link\]](#). The light is refracted both as it enters and as it leaves the drop. Since the index of refraction of water

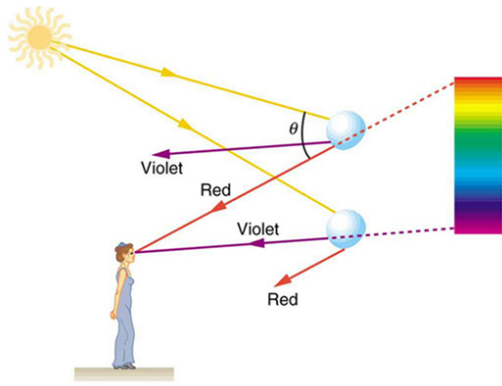
varies with wavelength, the light is dispersed, and a rainbow is observed, as shown in [\[link\]](#) (a). (There is no dispersion caused by reflection at the back surface, since the law of reflection does not depend on wavelength.) The actual rainbow of colors seen by an observer depends on the myriad of rays being refracted and reflected toward the observer's eyes from numerous drops of water. The effect is most spectacular when the background is dark, as in stormy weather, but can also be observed in waterfalls and lawn sprinklers. The arc of a rainbow comes from the need to be looking at a specific angle relative to the direction of the sun, as illustrated in [\[link\]](#) (b). (If there are two reflections of light within the water drop, another “secondary” rainbow is produced. This rare event produces an arc that lies above the primary rainbow arc—see [\[link\]](#) (c).)

Note:**Rainbows**

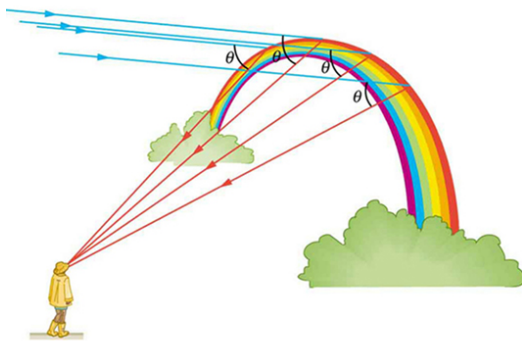
Rainbows are produced by a combination of refraction and reflection.



Part of the light falling on this water drop enters and is reflected from the back of the drop. This light is refracted and dispersed both as it enters and as it leaves the drop.



(a)



(b)



(c)

(a) Different colors emerge in different directions, and so you must look at different locations to see the various colors of a rainbow. (b) The arc of a rainbow results from the fact that a line between the observer and any point on the arc must make the correct angle with the parallel rays of sunlight to receive the refracted rays. (c)

Double rainbow. (credit:
Nicholas, Wikimedia
Commons)

Dispersion may produce beautiful rainbows, but it can cause problems in optical systems. White light used to transmit messages in a fiber is dispersed, spreading out in time and eventually overlapping with other messages. Since a laser produces a nearly pure wavelength, its light experiences little dispersion, an advantage over white light for transmission of information. In contrast, dispersion of electromagnetic waves coming to us from outer space can be used to determine the amount of matter they pass through. As with many phenomena, dispersion can be useful or a nuisance, depending on the situation and our human goals.

Note:

PhET Explorations: Geometric Optics

How does a lens form an image? See how light rays are refracted by a lens. Watch how the image changes when you adjust the focal length of the lens, move the object, move the lens, or move the screen.

https://phet.colorado.edu/sims/geometric-optics/geometric-optics_en.html

Section Summary

- The spreading of white light into its full spectrum of wavelengths is called dispersion.
- Rainbows are produced by a combination of refraction and reflection and involve the dispersion of sunlight into a continuous distribution of colors.
- Dispersion produces beautiful rainbows but also causes problems in certain optical systems.

Problems & Exercises

Exercise:

Problem:

- (a) What is the ratio of the speed of red light to violet light in diamond, based on [\[link\]](#)? (b) What is this ratio in polystyrene? (c) Which is more dispersive?

Exercise:**Problem:**

A beam of white light goes from air into water at an incident angle of 75.0° . At what angles are the red (660 nm) and violet (410 nm) parts of the light refracted?

Solution:

46.5° , red; 46.0° , violet

Exercise:**Problem:**

By how much do the critical angles for red (660 nm) and violet (410 nm) light differ in a diamond surrounded by air?

Exercise:**Problem:**

(a) A narrow beam of light containing yellow (580 nm) and green (550 nm) wavelengths goes from polystyrene to air, striking the surface at a 30.0° incident angle. What is the angle between the colors when they emerge? (b) How far would they have to travel to be separated by 1.00 mm?

Solution:

(a) 0.043°

(b) 1.33 m

Exercise:**Problem:**

A parallel beam of light containing orange (610 nm) and violet (410 nm) wavelengths goes from fused quartz to water, striking the surface between them at a 60.0° incident angle. What is the angle between the two colors in water?

Exercise:**Problem:**

A ray of 610 nm light goes from air into fused quartz at an incident angle of 55.0° . At what incident angle must 470 nm light enter flint glass to have the same angle of refraction?

Solution:

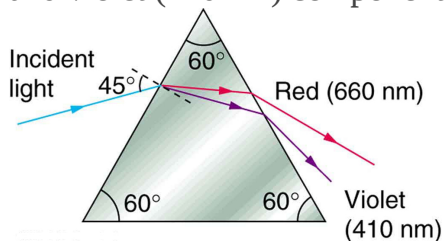
71.3°

Exercise:**Problem:**

A narrow beam of light containing red (660 nm) and blue (470 nm) wavelengths travels from air through a 1.00 cm thick flat piece of crown glass and back to air again. The beam strikes at a 30.0° incident angle. (a) At what angles do the two colors emerge? (b) By what distance are the red and blue separated when they emerge?

Exercise:**Problem:**

A narrow beam of white light enters a prism made of crown glass at a 45.0° incident angle, as shown in [\[link\]](#). At what angles, θ_R and θ_V , do the red (660 nm) and violet (410 nm) components of the light emerge from the prism?



This prism will disperse the white light into a rainbow of colors. The incident angle is 45.0°, and the angles at which the red and violet light emerge are θ_R and θ_V .

Solution:

53.5°, red; 55.2°, violet

Glossary

dispersion

spreading of white light into its full spectrum of wavelengths

rainbow

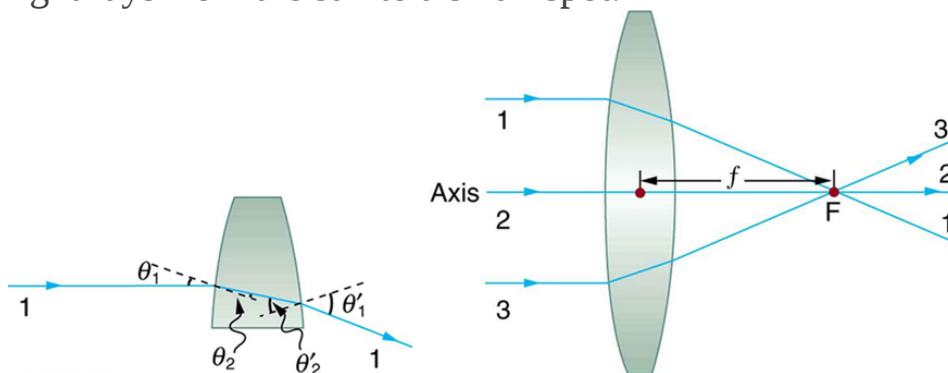
dispersion of sunlight into a continuous distribution of colors according to wavelength, produced by the refraction and reflection of sunlight by water droplets in the sky

Image Formation by Lenses

- List the rules for ray tracking for thin lenses.
- Illustrate the formation of images using the technique of ray tracking.
- Determine power of a lens given the focal length.

Lenses are found in a huge array of optical instruments, ranging from a simple magnifying glass to the eye to a camera's zoom lens. In this section, we will use the law of refraction to explore the properties of lenses and how they form images.

The word *lens* derives from the Latin word for a lentil bean, the shape of which is similar to the convex lens in [\[link\]](#). The convex lens shown has been shaped so that all light rays that enter it parallel to its axis cross one another at a single point on the opposite side of the lens. (The axis is defined to be a line normal to the lens at its center, as shown in [\[link\]](#).) Such a lens is called a **converging (or convex) lens** for the converging effect it has on light rays. An expanded view of the path of one ray through the lens is shown, to illustrate how the ray changes direction both as it enters and as it leaves the lens. Since the index of refraction of the lens is greater than that of air, the ray moves towards the perpendicular as it enters and away from the perpendicular as it leaves. (This is in accordance with the law of refraction.) Due to the lens's shape, light is thus bent toward the axis at both surfaces. The point at which the rays cross is defined to be the **focal point F** of the lens. The distance from the center of the lens to its focal point is defined to be the **focal length f** of the lens. [\[link\]](#) shows how a converging lens, such as that in a magnifying glass, can converge the nearly parallel light rays from the sun to a small spot.



Rays of light entering a converging lens parallel to its axis converge at its focal point F. (Ray 2 lies on the axis of the lens.) The distance from the center of the lens to the focal point is the lens's focal length f . An expanded view of the path taken by ray 1 shows the perpendiculars and the angles of incidence and refraction at both surfaces.

Note:

Converging or Convex Lens

The lens in which light rays that enter it parallel to its axis cross one another at a single point on the opposite side with a converging effect is called converging lens.

Note:

Focal Point F

The point at which the light rays cross is called the focal point F of the lens.

Note:

Focal Length f

The distance from the center of the lens to its focal point is called focal length f .



Sunlight focused by a converging magnifying glass can burn paper. Light rays from the sun are nearly parallel and cross at the focal point of the lens. The more powerful the lens, the closer to the lens the rays will cross.

The greater effect a lens has on light rays, the more powerful it is said to be. For example, a powerful converging lens will focus parallel light rays closer to itself and will have a smaller focal length than a weak lens. The light will also focus into a smaller and more intense spot for a more powerful lens. The **power** P of a lens is defined to be the inverse of its focal length. In equation form, this is

Equation:

$$P = \frac{1}{f}.$$

Note:**Power P**

The **power** P of a lens is defined to be the inverse of its focal length. In equation form, this is

Equation:

$$P = \frac{1}{f}.$$

where f is the focal length of the lens, which must be given in meters (and not cm or mm). The power of a lens P has the unit diopters (D), provided that the focal length is given in meters. That is, $1 \text{ D} = 1/\text{m}$, or 1 m^{-1} .

(Note that this power (optical power, actually) is not the same as power in watts defined in [Work, Energy, and Energy Resources](#). It is a concept related to the effect of optical devices on light.) Optometrists prescribe common spectacles and contact lenses in units of diopters.

Example:**What is the Power of a Common Magnifying Glass?**

Suppose you take a magnifying glass out on a sunny day and you find that it concentrates sunlight to a small spot 8.00 cm away from the lens. What are the focal length and power of the lens?

Strategy

The situation here is the same as those shown in [\[link\]](#) and [\[link\]](#). The Sun is so far away that the Sun's rays are nearly parallel when they reach Earth. The magnifying glass is a convex (or converging) lens, focusing the nearly parallel rays of sunlight. Thus the focal length of the lens is the distance from the lens to the spot, and its power is the inverse of this distance (in m).

Solution

The focal length of the lens is the distance from the center of the lens to the spot, given to be 8.00 cm. Thus,

Equation:

$$f = 8.00 \text{ cm}.$$

To find the power of the lens, we must first convert the focal length to meters; then, we substitute this value into the equation for power. This gives

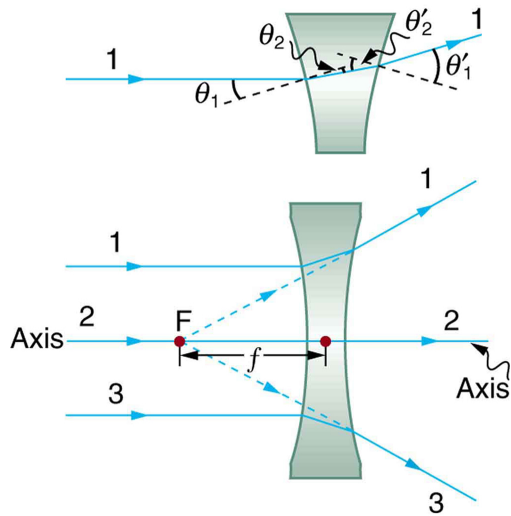
Equation:

$$P = \frac{1}{f} = \frac{1}{0.0800 \text{ m}} = 12.5 \text{ D.}$$

Discussion

This is a relatively powerful lens. The power of a lens in diopters should not be confused with the familiar concept of power in watts. It is an unfortunate fact that the word “power” is used for two completely different concepts. If you examine a prescription for eyeglasses, you will note lens powers given in diopters. If you examine the label on a motor, you will note energy consumption rate given as a power in watts.

[\[link\]](#) shows a concave lens and the effect it has on rays of light that enter it parallel to its axis (the path taken by ray 2 in the figure is the axis of the lens). The concave lens is a **diverging lens**, because it causes the light rays to bend away (diverge) from its axis. In this case, the lens has been shaped so that all light rays entering it parallel to its axis appear to originate from the same point, F , defined to be the focal point of a diverging lens. The distance from the center of the lens to the focal point is again called the focal length f of the lens. Note that the focal length and power of a diverging lens are defined to be negative. For example, if the distance to F in [\[link\]](#) is 5.00 cm, then the focal length is $f = -5.00 \text{ cm}$ and the power of the lens is $P = -20 \text{ D}$. An expanded view of the path of one ray through the lens is shown in the figure to illustrate how the shape of the lens, together with the law of refraction, causes the ray to follow its particular path and be diverged.

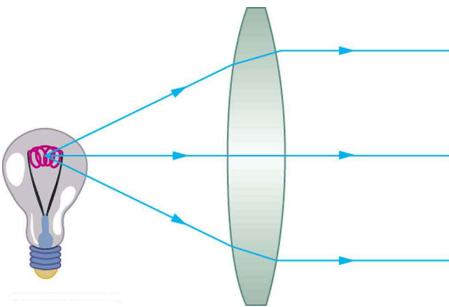


Rays of light entering a diverging lens parallel to its axis are diverged, and all appear to originate at its focal point F. The dashed lines are not rays—they indicate the directions from which the rays appear to come. The focal length f of a diverging lens is negative. An expanded view of the path taken by ray 1 shows the perpendiculars and the angles of incidence and refraction at both surfaces.

Note:
Diverging Lens

A lens that causes the light rays to bend away from its axis is called a diverging lens.

As noted in the initial discussion of the law of refraction in [The Law of Refraction](#), the paths of light rays are exactly reversible. This means that the direction of the arrows could be reversed for all of the rays in [\[link\]](#) and [\[link\]](#). For example, if a point light source is placed at the focal point of a convex lens, as shown in [\[link\]](#), parallel light rays emerge from the other side.



A small light source, like a light bulb filament, placed at the focal point of a convex lens, results in parallel rays of light emerging from the other side. The paths are exactly the reverse of those shown in [\[link\]](#). This technique is used in lighthouses and sometimes in traffic lights to produce a directional beam of light from a source that emits light in all directions.

Ray Tracing and Thin Lenses

Ray tracing is the technique of determining or following (tracing) the paths that light rays take. For rays passing through matter, the law of refraction is used to trace the paths. Here we use ray tracing to help us understand the action of lenses in situations ranging from forming images on film to magnifying small print to correcting nearsightedness. While ray tracing for complicated lenses, such as those found in sophisticated cameras, may require computer techniques, there is a set of simple rules for tracing rays through thin lenses. A **thin lens** is defined to be one whose thickness allows rays to refract, as illustrated in [\[link\]](#), but does not allow properties such as dispersion and aberrations. An ideal thin lens has two refracting surfaces but the lens is thin enough to assume that light rays bend only once. A thin symmetrical lens has two focal points, one on either side and both at the same distance from the lens. (See [\[link\]](#).) Another important characteristic of a thin lens is that light rays through its center are deflected by a negligible amount, as seen in [\[link\]](#).

Note:

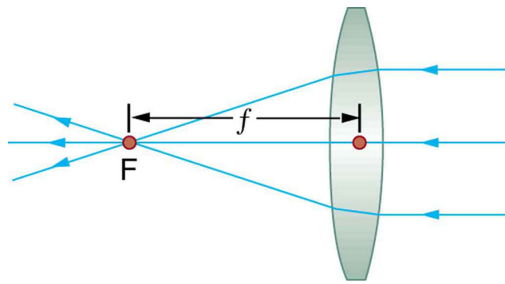
Thin Lens

A thin lens is defined to be one whose thickness allows rays to refract but does not allow properties such as dispersion and aberrations.

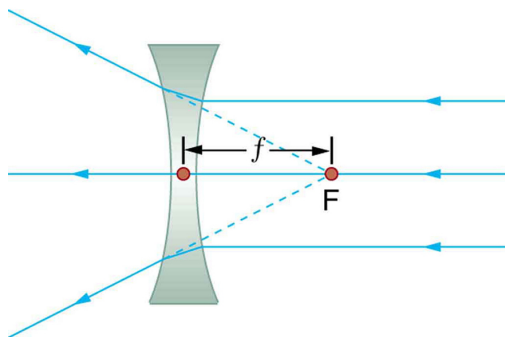
Note:

Take-Home Experiment: A Visit to the Optician

Look through your eyeglasses (or those of a friend) backward and forward and comment on whether they act like thin lenses.

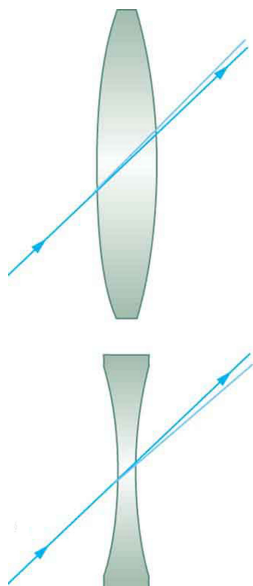


(a)



(b)

Thin lenses have the same focal length on either side. (a) Parallel light rays entering a converging lens from the right cross at its focal point on the left. (b) Parallel light rays entering a diverging lens from the right seem to come from the focal point on the right.



The light ray through the center of a thin lens is deflected by a negligible amount and is assumed to emerge parallel to its original path (shown as a shaded line).

Using paper, pencil, and a straight edge, ray tracing can accurately describe the operation of a lens. The rules for ray tracing for thin lenses are based on the illustrations already discussed:

1. A ray entering a converging lens parallel to its axis passes through the focal point F of the lens on the other side. (See rays 1 and 3 in [\[link\]](#).)
2. A ray entering a diverging lens parallel to its axis seems to come from the focal point F . (See rays 1 and 3 in [\[link\]](#).)
3. A ray passing through the center of either a converging or a diverging lens does not change direction. (See [\[link\]](#), and see ray 2 in [\[link\]](#) and [\[link\]](#).)
4. A ray entering a converging lens through its focal point exits parallel to its axis. (The reverse of rays 1 and 3 in [\[link\]](#).)
5. A ray that enters a diverging lens by heading toward the focal point on the opposite side exits parallel to the axis. (The reverse of rays 1 and 3 in [\[link\]](#).)

Note:

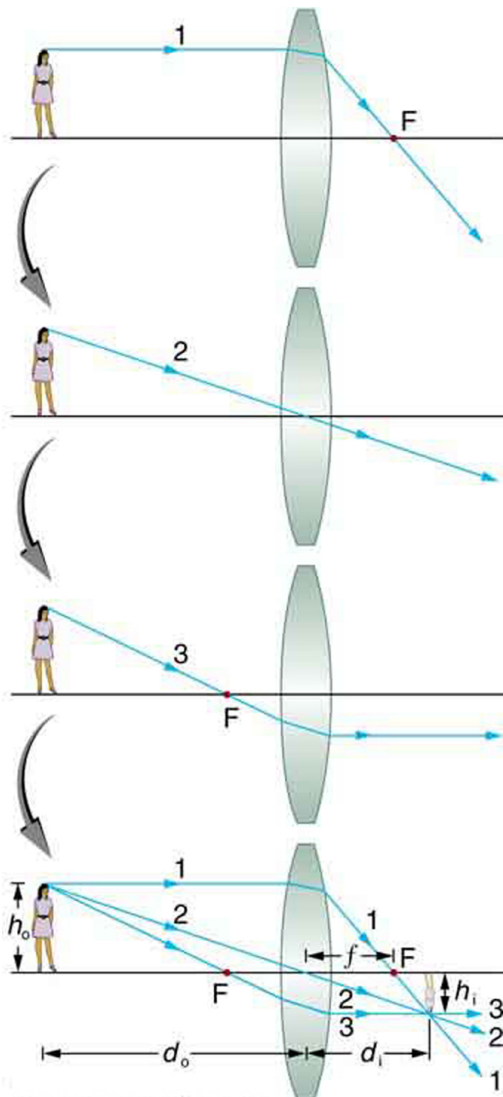
Rules for Ray Tracing

1. A ray entering a converging lens parallel to its axis passes through the focal point F of the lens on the other side.
2. A ray entering a diverging lens parallel to its axis seems to come from the focal point F .
3. A ray passing through the center of either a converging or a diverging lens does not change direction.
4. A ray entering a converging lens through its focal point exits parallel to its axis.
5. A ray that enters a diverging lens by heading toward the focal point on the opposite side exits parallel to the axis.

Image Formation by Thin Lenses

In some circumstances, a lens forms an obvious image, such as when a movie projector casts an image onto a screen. In other cases, the image is less obvious. Where, for example, is the image formed by eyeglasses? We will use ray tracing for thin lenses to illustrate how they form images, and we will develop equations to describe the image formation quantitatively.

Consider an object some distance away from a converging lens, as shown in [\[link\]](#). To find the location and size of the image formed, we trace the paths of selected light rays originating from one point on the object, in this case the top of the person's head. The figure shows three rays from the top of the object that can be traced using the ray tracing rules given above. (Rays leave this point going in many directions, but we concentrate on only a few with paths that are easy to trace.) The first ray is one that enters the lens parallel to its axis and passes through the focal point on the other side (rule 1). The second ray passes through the center of the lens without changing direction (rule 3). The third ray passes through the nearer focal point on its way into the lens and leaves the lens parallel to its axis (rule 4). The three rays cross at the same point on the other side of the lens. The image of the top of the person's head is located at this point. All rays that come from the same point on the top of the person's head are refracted in such a way as to cross at the point shown. Rays from another point on the object, such as her belt buckle, will also cross at another common point, forming a complete image, as shown. Although three rays are traced in [\[link\]](#), only two are necessary to locate the image. It is best to trace rays for which there are simple ray tracing rules. Before applying ray tracing to other situations, let us consider the example shown in [\[link\]](#) in more detail.



Ray tracing is used to locate the image formed by a lens. Rays originating from the same point on the object are traced—the three chosen rays each follow one of the rules for ray tracing, so that their paths are easy to determine. The image is located at the point where the rays

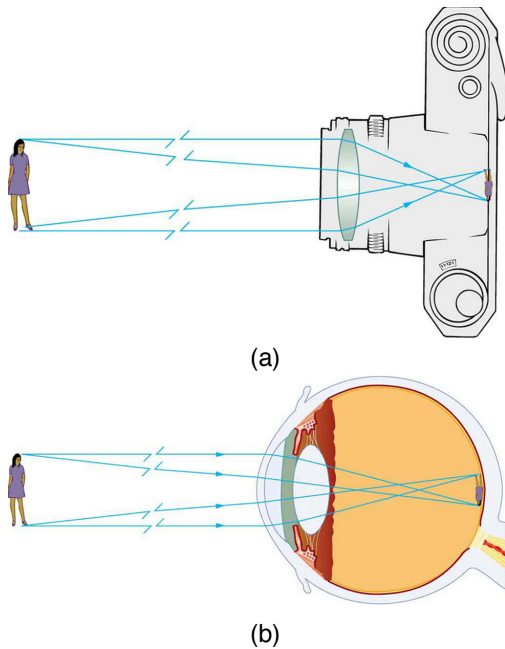
cross. In this case, a real image—one that can be projected on a screen—is formed.

The image formed in [\[link\]](#) is a **real image**, meaning that it can be projected. That is, light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye, for example. [\[link\]](#) shows how such an image would be projected onto film by a camera lens. This figure also shows how a real image is projected onto the retina by the lens of an eye. Note that the image is there whether it is projected onto a screen or not.

Note:

Real Image

The image in which light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye is called a real image.



Real images can be projected. (a) A real image of the person is projected onto film. (b) The converging nature of the multiple surfaces that make up the eye result in the projection of a real image on the retina.

Several important distances appear in [\[link\]](#). We define d_o to be the object distance, the distance of an object from the center of a lens. **Image distance** d_i is defined to be the distance of the image from the center of a lens. The height of the object and height of the image are given the symbols h_o and h_i , respectively. Images that appear upright relative to the object have heights that are positive and those that are inverted have negative heights. Using the rules of ray tracing and making a scale drawing with paper and pencil, like that in [\[link\]](#), we can accurately describe the location and size of an image. But the real benefit of ray tracing is in visualizing how images are formed in a variety of situations. To obtain numerical information, we use a pair of

equations that can be derived from a geometric analysis of ray tracing for thin lenses. The **thin lens equations** are

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

and

Equation:

$$\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m.$$

We define the ratio of image height to object height (h_i/h_o) to be the **magnification** m . (The minus sign in the equation above will be discussed shortly.) The thin lens equations are broadly applicable to all situations involving thin lenses (and “thin” mirrors, as we will see later). We will explore many features of image formation in the following worked examples.

Note:

Image Distance

The distance of the image from the center of the lens is called image distance.

Note:

Thin Lens Equations and Magnification

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

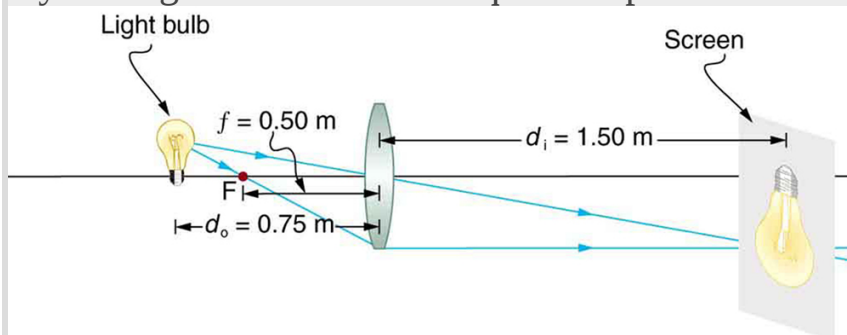
Equation:

$$\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m$$

Example:

Finding the Image of a Light Bulb Filament by Ray Tracing and by the Thin Lens Equations

A clear glass light bulb is placed 0.750 m from a convex lens having a 0.500 m focal length, as shown in [\[link\]](#). Use ray tracing to get an approximate location for the image. Then use the thin lens equations to calculate (a) the location of the image and (b) its magnification. Verify that ray tracing and the thin lens equations produce consistent results.



A light bulb placed 0.750 m from a lens having a 0.500 m focal length produces a real image on a poster board as discussed in the example above. Ray tracing predicts the image location and size.

Strategy and Concept

Since the object is placed farther away from a converging lens than the focal length of the lens, this situation is analogous to those illustrated in [\[link\]](#) and [\[link\]](#). Ray tracing to scale should produce similar results for d_i . Numerical solutions for d_i and m can be obtained using the thin lens equations, noting that $d_o = 0.750 \text{ m}$ and $f = 0.500 \text{ m}$.

Solutions (Ray tracing)

The ray tracing to scale in [\[link\]](#) shows two rays from a point on the bulb's filament crossing about 1.50 m on the far side of the lens. Thus the image distance d_i is about 1.50 m. Similarly, the image height based on ray tracing is greater than the object height by about a factor of 2, and the image is inverted. Thus m is about -2 . The minus sign indicates that the image is inverted.

The thin lens equations can be used to find d_i from the given information:

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}.$$

Rearranging to isolate d_i gives

Equation:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o}.$$

Entering known quantities gives a value for $1/d_i$:

Equation:

$$\frac{1}{d_i} = \frac{1}{0.500 \text{ m}} - \frac{1}{0.750 \text{ m}} = \frac{0.667}{\text{m}}.$$

This must be inverted to find d_i :

Equation:

$$d_i = \frac{\text{m}}{0.667} = 1.50 \text{ m}.$$

Note that another way to find d_i is to rearrange the equation:

Equation:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o}.$$

This yields the equation for the image distance as:

Equation:

$$d_i = \frac{fd_o}{d_o - f}.$$

Note that there is no inverting here.

The thin lens equations can be used to find the magnification m , since both d_i and d_o are known. Entering their values gives

Equation:

$$m = -\frac{d_i}{d_o} = -\frac{1.50 \text{ m}}{0.750 \text{ m}} = -2.00.$$

Discussion

Note that the minus sign causes the magnification to be negative when the image is inverted. Ray tracing and the use of the thin lens equations produce consistent results. The thin lens equations give the most precise results, being limited only by the accuracy of the given information. Ray tracing is limited by the accuracy with which you can draw, but it is highly useful both conceptually and visually.

Real images, such as the one considered in the previous example, are formed by converging lenses whenever an object is farther from the lens than its focal length. This is true for movie projectors, cameras, and the eye. We shall refer to these as *case 1* images. A case 1 image is formed when $d_o > f$ and f is positive, as in [\[link\]](#)(a). (A summary of the three cases or types of image formation appears at the end of this section.)

A different type of image is formed when an object, such as a person's face, is held close to a convex lens. The image is upright and larger than the object, as seen in [\[link\]](#)(b), and so the lens is called a magnifier. If you slowly pull the magnifier away from the face, you will see that the magnification steadily increases until the image begins to blur. Pulling the magnifier even farther away produces an inverted image as seen in [\[link\]](#)(a). The distance at which the image blurs, and beyond which it inverts, is the focal length of the lens. To use a convex lens as a magnifier, the object

must be closer to the converging lens than its focal length. This is called a *case 2* image. A case 2 image is formed when $d_o < f$ and f is positive.



(a)



(b)

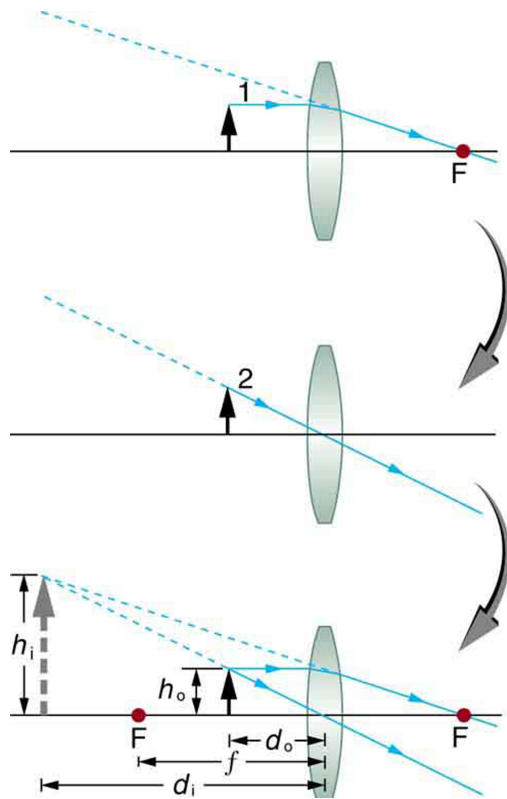
(a) When a converging lens is held farther away from the face than the lens's focal length, an inverted image is formed. This is a case 1 image. Note that the image is in focus but the face is not, because the image is much closer to the camera taking this photograph than the face. (credit:

DaMongMan, Flickr)

(b) A magnified image

of a face is produced
by placing it closer to
the converging lens
than its focal length.
This is a case 2 image.
(credit: Casey Fleser,
Flickr)

[\[link\]](#) uses ray tracing to show how an image is formed when an object is held closer to a converging lens than its focal length. Rays coming from a common point on the object continue to diverge after passing through the lens, but all appear to originate from a point at the location of the image. The image is on the same side of the lens as the object and is farther away from the lens than the object. This image, like all case 2 images, cannot be projected and, hence, is called a **virtual image**. Light rays only appear to originate at a virtual image; they do not actually pass through that location in space. A screen placed at the location of a virtual image will receive only diffuse light from the object, not focused rays from the lens. Additionally, a screen placed on the opposite side of the lens will receive rays that are still diverging, and so no image will be projected on it. We can see the magnified image with our eyes, because the lens of the eye converges the rays into a real image projected on our retina. Finally, we note that a virtual image is upright and larger than the object, meaning that the magnification is positive and greater than 1.



Ray tracing predicts the image location and size for an object held closer to a converging lens than its focal length. Ray 1 enters parallel to the axis and exits through the focal point on the opposite side, while ray 2 passes through the center of the lens without changing path. The two rays continue to diverge on the other side of the lens, but both appear to come from a common point, locating the upright, magnified,

virtual image. This is a case 2 image.

Note:

Virtual Image

An image that is on the same side of the lens as the object and cannot be projected on a screen is called a virtual image.

Example:

Image Produced by a Magnifying Glass

Suppose the book page in [\[link\]](#) (a) is held 7.50 cm from a convex lens of focal length 10.0 cm, such as a typical magnifying glass might have. What magnification is produced?

Strategy and Concept

We are given that $d_o = 7.50$ cm and $f = 10.0$ cm, so we have a situation where the object is placed closer to the lens than its focal length. We therefore expect to get a case 2 virtual image with a positive magnification that is greater than 1. Ray tracing produces an image like that shown in [\[link\]](#), but we will use the thin lens equations to get numerical solutions in this example.

Solution

To find the magnification m , we try to use magnification equation, $m = -d_i/d_o$. We do not have a value for d_i , so that we must first find the location of the image using lens equation. (The procedure is the same as followed in the preceding example, where d_o and f were known.)

Rearranging the magnification equation to isolate d_i gives

Equation:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o}.$$

Entering known values, we obtain a value for $1/d_i$:

Equation:

$$\frac{1}{d_i} = \frac{1}{10.0 \text{ cm}} - \frac{1}{7.50 \text{ cm}} = \frac{-0.0333}{\text{cm}}.$$

This must be inverted to find d_i :

Equation:

$$d_i = -\frac{\text{cm}}{0.0333} = -30.0 \text{ cm}.$$

Now the thin lens equation can be used to find the magnification m , since both d_i and d_o are known. Entering their values gives

Equation:

$$m = -\frac{d_i}{d_o} = -\frac{-30.0 \text{ cm}}{7.50 \text{ cm}} = 4.00.$$

Discussion

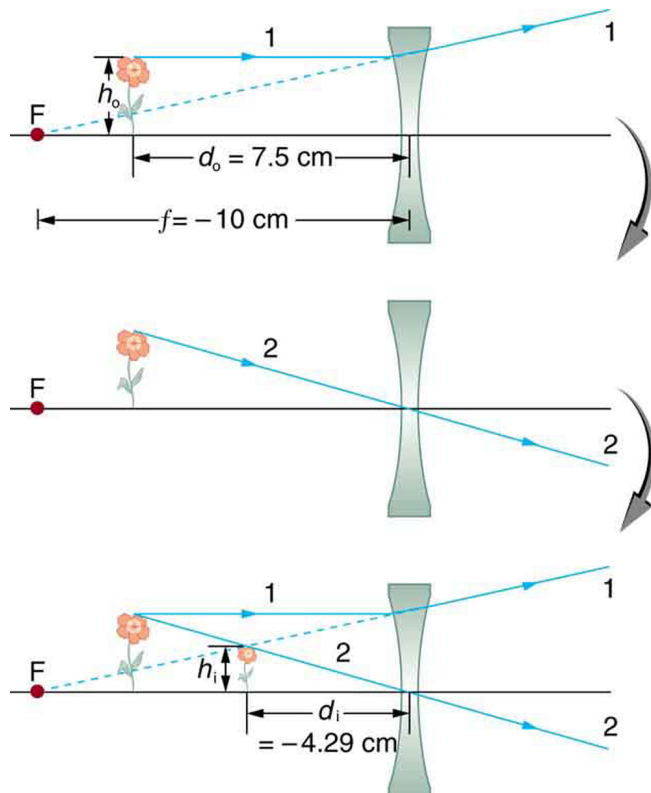
A number of results in this example are true of all case 2 images, as well as being consistent with [\[link\]](#). Magnification is indeed positive (as predicted), meaning the image is upright. The magnification is also greater than 1, meaning that the image is larger than the object—in this case, by a factor of 4. Note that the image distance is negative. This means the image is on the same side of the lens as the object. Thus the image cannot be projected and is virtual. (Negative values of d_i occur for virtual images.) The image is farther from the lens than the object, since the image distance is greater in magnitude than the object distance. The location of the image is not obvious when you look through a magnifier. In fact, since the image is bigger than the object, you may think the image is closer than the object. But the image is farther away, a fact that is useful in correcting farsightedness, as we shall see in a later section.

A third type of image is formed by a diverging or concave lens. Try looking through eyeglasses meant to correct nearsightedness. (See [\[link\]](#).) You will see an image that is upright but smaller than the object. This means that the magnification is positive but less than 1. The ray diagram in [\[link\]](#) shows that the image is on the same side of the lens as the object and, hence,

cannot be projected—it is a virtual image. Note that the image is closer to the lens than the object. This is a *case 3* image, formed for any object by a negative focal length or diverging lens.



A car viewed through a concave or diverging lens looks upright. This is a case 3 image. (credit: Daniel Oines, Flickr)



Ray tracing predicts the image location and size for a concave or diverging lens. Ray 1 enters parallel to the axis and is bent so that it appears to originate from the focal point. Ray 2 passes through the center of the lens without changing path. The two rays appear to come from a common point, locating the upright image. This is a case 3 image, which is closer to the lens than the object and smaller in height.

Example:

Image Produced by a Concave Lens

Suppose an object such as a book page is held 7.50 cm from a concave lens of focal length -10.0 cm. Such a lens could be used in eyeglasses to correct pronounced nearsightedness. What magnification is produced?

Strategy and Concept

This example is identical to the preceding one, except that the focal length is negative for a concave or diverging lens. The method of solution is thus the same, but the results are different in important ways.

Solution

To find the magnification m , we must first find the image distance d_i using thin lens equation

Equation:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o},$$

or its alternative rearrangement

Equation:

$$d_i = \frac{fd_o}{d_o - f}.$$

We are given that $f = -10.0$ cm and $d_o = 7.50$ cm. Entering these yields a value for $1/d_i$:

Equation:

$$\frac{1}{d_i} = \frac{1}{-10.0 \text{ cm}} - \frac{1}{7.50 \text{ cm}} = \frac{-0.2333}{\text{cm}}.$$

This must be inverted to find d_i :

Equation:

$$d_i = -\frac{\text{cm}}{0.2333} = -4.29 \text{ cm}.$$

Or

Equation:

$$d_i = \frac{(7.5)(-10)}{(7.5 - (-10))} = -75/17.5 = -4.29 \text{ cm.}$$

Now the magnification equation can be used to find the magnification m , since both d_i and d_o are known. Entering their values gives

Equation:

$$m = -\frac{d_i}{d_o} = -\frac{-4.29 \text{ cm}}{7.50 \text{ cm}} = 0.571.$$

Discussion

A number of results in this example are true of all case 3 images, as well as being consistent with [\[link\]](#). Magnification is positive (as predicted), meaning the image is upright. The magnification is also less than 1, meaning the image is smaller than the object—in this case, a little over half its size. The image distance is negative, meaning the image is on the same side of the lens as the object. (The image is virtual.) The image is closer to the lens than the object, since the image distance is smaller in magnitude than the object distance. The location of the image is not obvious when you look through a concave lens. In fact, since the image is smaller than the object, you may think it is farther away. But the image is closer than the object, a fact that is useful in correcting nearsightedness, as we shall see in a later section.

[\[link\]](#) summarizes the three types of images formed by single thin lenses. These are referred to as case 1, 2, and 3 images. Convex (converging) lenses can form either real or virtual images (cases 1 and 2, respectively), whereas concave (diverging) lenses can form only virtual images (always case 3). Real images are always inverted, but they can be either larger or smaller than the object. For example, a slide projector forms an image larger than the slide, whereas a camera makes an image smaller than the object being photographed. Virtual images are always upright and cannot be projected. Virtual images are larger than the object only in case 2, where a convex lens is used. The virtual image produced by a concave lens is

always smaller than the object—a case 3 image. We can see and photograph virtual images only by using an additional lens to form a real image.

Type	Formed when	Image type	d_i	m
Case 1	f positive, $d_o > f$	real	positive	negative
Case 2	f positive, $d_o < f$	virtual	negative	positive $m > 1$
Case 3	f negative	virtual	negative	positive $m < 1$

Three Types of Images Formed By Thin Lenses

In [Image Formation by Mirrors](#), we shall see that mirrors can form exactly the same types of images as lenses.

Note:

Take-Home Experiment: Concentrating Sunlight

Find several lenses and determine whether they are converging or diverging. In general those that are thicker near the edges are diverging and those that are thicker near the center are converging. On a bright sunny day take the converging lenses outside and try focusing the sunlight onto a piece of paper. Determine the focal lengths of the lenses. Be careful because the paper may start to burn, depending on the type of lens you have selected.

Problem-Solving Strategies for Lenses

Step 1. Examine the situation to determine that image formation by a lens is involved.

Step 2. Determine whether ray tracing, the thin lens equations, or both are to be employed. A sketch is very useful even if ray tracing is not specifically required by the problem. Write symbols and values on the sketch.

Step 3. Identify exactly what needs to be determined in the problem (identify the unknowns).

Step 4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns). It is helpful to determine whether the situation involves a case 1, 2, or 3 image. While these are just names for types of images, they have certain characteristics (given in [\[link\]](#)) that can be of great use in solving problems.

Step 5. If ray tracing is required, use the ray tracing rules listed near the beginning of this section.

Step 6. Most quantitative problems require the use of the thin lens equations. These are solved in the usual manner by substituting knowns and solving for unknowns. Several worked examples serve as guides.

Step 7. Check to see if the answer is reasonable: Does it make sense? If you have identified the type of image (case 1, 2, or 3), you should assess whether your answer is consistent with the type of image, magnification, and so on.

Note:

Misconception Alert

We do not realize that light rays are coming from every part of the object, passing through every part of the lens, and all can be used to form the final image.

We generally feel the entire lens, or mirror, is needed to form an image. Actually, half a lens will form the same, though a fainter, image.

Section Summary

- Light rays entering a converging lens parallel to its axis cross one another at a single point on the opposite side.
- For a converging lens, the focal point is the point at which converging light rays cross; for a diverging lens, the focal point is the point from which diverging light rays appear to originate.
- The distance from the center of the lens to its focal point is called the focal length f .
- Power P of a lens is defined to be the inverse of its focal length, $P = \frac{1}{f}$.
- A lens that causes the light rays to bend away from its axis is called a diverging lens.
- Ray tracing is the technique of graphically determining the paths that light rays take.
- The image in which light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye is called a real image.
- Thin lens equations are $\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$ and $\frac{h_i}{h_o} = -\frac{d_i}{d_o} = m$ (magnification).

- The distance of the image from the center of the lens is called image distance.
- An image that is on the same side of the lens as the object and cannot be projected on a screen is called a virtual image.

Conceptual Questions

Exercise:

Problem:

It can be argued that a flat piece of glass, such as in a window, is like a lens with an infinite focal length. If so, where does it form an image? That is, how are d_i and d_o related?

Exercise:

Problem:

You can often see a reflection when looking at a sheet of glass, particularly if it is darker on the other side. Explain why you can often see a double image in such circumstances.

Exercise:

Problem:

When you focus a camera, you adjust the distance of the lens from the film. If the camera lens acts like a thin lens, why can it not be a fixed distance from the film for both near and distant objects?

Exercise:

Problem:

A thin lens has two focal points, one on either side, at equal distances from its center, and should behave the same for light entering from either side. Look through your eyeglasses (or those of a friend) backward and forward and comment on whether they are thin lenses.

Exercise:

Problem:

Will the focal length of a lens change when it is submerged in water? Explain.

Problems & Exercises**Exercise:****Problem:**

What is the power in diopters of a camera lens that has a 50.0 mm focal length?

Exercise:**Problem:**

Your camera's zoom lens has an adjustable focal length ranging from 80.0 to 200 mm. What is its range of powers?

Solution:

5.00 to 12.5 D

Exercise:**Problem:**

What is the focal length of 1.75 D reading glasses found on the rack in a pharmacy?

Exercise:**Problem:**

You note that your prescription for new eyeglasses is -4.50 D. What will their focal length be?

Solution:

-0.222 m

Exercise:

Problem:

How far from the lens must the film in a camera be, if the lens has a 35.0 mm focal length and is being used to photograph a flower 75.0 cm away? Explicitly show how you follow the steps in the Problem-Solving Strategy for lenses.

Exercise:

Problem:

A certain slide projector has a 100 mm focal length lens. (a) How far away is the screen, if a slide is placed 103 mm from the lens and produces a sharp image? (b) If the slide is 24.0 by 36.0 mm, what are the dimensions of the image? Explicitly show how you follow the steps in the Problem-Solving Strategy for lenses.

Solution:

(a) 3.43 m

(b) 0.800 by 1.20 m

Exercise:

Problem:

A doctor examines a mole with a 15.0 cm focal length magnifying glass held 13.5 cm from the mole (a) Where is the image? (b) What is its magnification? (c) How big is the image of a 5.00 mm diameter mole?

Solution:

(a) -1.35 m (on the object side of the lens).

(b) $+10.0$

(c) 5.00 cm

Exercise:

Problem:

How far from a piece of paper must you hold your father's 2.25 D reading glasses to try to burn a hole in the paper with sunlight?

Solution:

44.4 cm

Exercise:

Problem:

A camera with a 50.0 mm focal length lens is being used to photograph a person standing 3.00 m away. (a) How far from the lens must the film be? (b) If the film is 36.0 mm high, what fraction of a 1.75 m tall person will fit on it? (c) Discuss how reasonable this seems, based on your experience in taking or posing for photographs.

Exercise:

Problem:

A camera lens used for taking close-up photographs has a focal length of 22.0 mm. The farthest it can be placed from the film is 33.0 mm. (a) What is the closest object that can be photographed? (b) What is the magnification of this closest object?

Solution:

(a) 6.60 cm

(b) -0.333

Exercise:

Problem:

Suppose your 50.0 mm focal length camera lens is 51.0 mm away from the film in the camera. (a) How far away is an object that is in focus? (b) What is the height of the object if its image is 2.00 cm high?

Exercise:**Problem:**

(a) What is the focal length of a magnifying glass that produces a magnification of 3.00 when held 5.00 cm from an object, such as a rare coin? (b) Calculate the power of the magnifier in diopters. (c) Discuss how this power compares to those for store-bought reading glasses (typically 1.0 to 4.0 D). Is the magnifier's power greater, and should it be?

Solution:

(a) +7.50 cm

(b) 13.3 D

(c) Much greater

Exercise:**Problem:**

What magnification will be produced by a lens of power -4.00 D (such as might be used to correct myopia) if an object is held 25.0 cm away?

Exercise:**Problem:**

In [\[link\]](#), the magnification of a book held 7.50 cm from a 10.0 cm focal length lens was found to be 3.00. (a) Find the magnification for the book when it is held 8.50 cm from the magnifier. (b) Do the same for when it is held 9.50 cm from the magnifier. (c) Comment on the trend in m as the object distance increases as in these two calculations.

Solution:

(a) +6.67

(b) +20.0

(c) The magnification increases without limit (to infinity) as the object distance increases to the limit of the focal distance.

Exercise:**Problem:**

Suppose a 200 mm focal length telephoto lens is being used to photograph mountains 10.0 km away. (a) Where is the image? (b) What is the height of the image of a 1000 m high cliff on one of the mountains?

Exercise:**Problem:**

A camera with a 100 mm focal length lens is used to photograph the sun and moon. What is the height of the image of the sun on the film, given the sun is 1.40×10^6 km in diameter and is 1.50×10^8 km away?

Solution:

−0.933 mm

Exercise:**Problem:**

Combine thin lens equations to show that the magnification for a thin lens is determined by its focal length and the object distance and is given by $m = f/(f - d_o)$.

Glossary

converging lens

a convex lens in which light rays that enter it parallel to its axis converge at a single point on the opposite side

diverging lens

a concave lens in which light rays that enter it parallel to its axis bend away (diverge) from its axis

focal point

for a converging lens or mirror, the point at which converging light rays cross; for a diverging lens or mirror, the point from which diverging light rays appear to originate

focal length

distance from the center of a lens or curved mirror to its focal point

magnification

ratio of image height to object height

power

inverse of focal length

real image

image that can be projected

virtual image

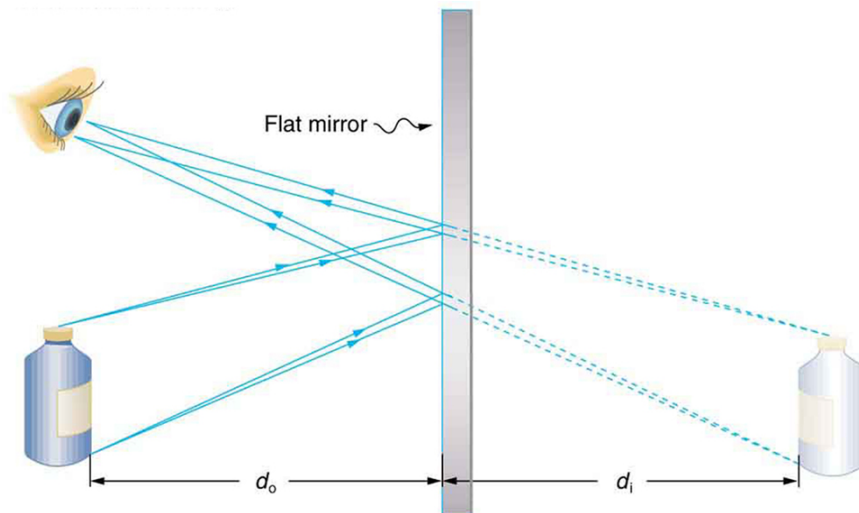
image that cannot be projected

Image Formation by Mirrors

- Illustrate image formation in a flat mirror.
- Explain with ray diagrams the formation of an image using spherical mirrors.
- Determine focal length and magnification given radius of curvature, distance of object and image.

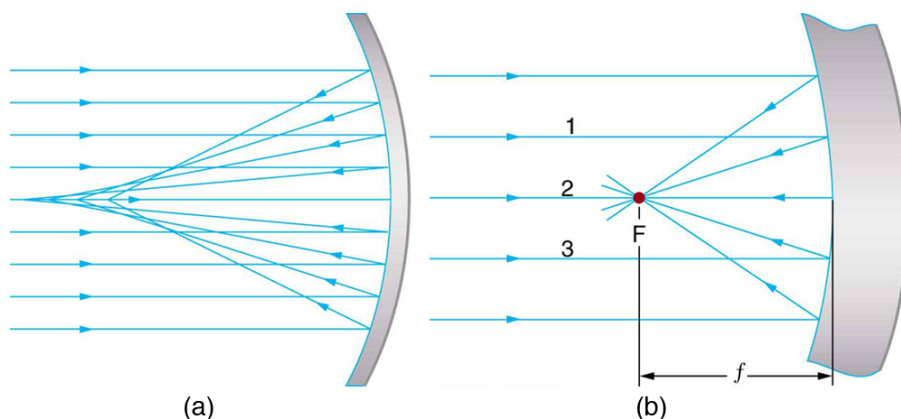
We only have to look as far as the nearest bathroom to find an example of an image formed by a mirror. Images in flat mirrors are the same size as the object and are located behind the mirror. Like lenses, mirrors can form a variety of images. For example, dental mirrors may produce a magnified image, just as makeup mirrors do. Security mirrors in shops, on the other hand, form images that are smaller than the object. We will use the law of reflection to understand how mirrors form images, and we will find that mirror images are analogous to those formed by lenses.

[\[link\]](#) helps illustrate how a flat mirror forms an image. Two rays are shown emerging from the same point, striking the mirror, and being reflected into the observer's eye. The rays can diverge slightly, and both still get into the eye. If the rays are extrapolated backward, they seem to originate from a common point behind the mirror, locating the image. (The paths of the reflected rays into the eye are the same as if they had come directly from that point behind the mirror.) Using the law of reflection—the angle of reflection equals the angle of incidence—we can see that the image and object are the same distance from the mirror. This is a virtual image, since it cannot be projected—the rays only appear to originate from a common point behind the mirror. Obviously, if you walk behind the mirror, you cannot see the image, since the rays do not go there. But in front of the mirror, the rays behave exactly as if they had come from behind the mirror, so that is where the image is situated.



Two sets of rays from common points on an object are reflected by a flat mirror into the eye of an observer. The reflected rays seem to originate from behind the mirror, locating the virtual image.

Now let us consider the focal length of a mirror—for example, the concave spherical mirrors in [\[link\]](#). Rays of light that strike the surface follow the law of reflection. For a mirror that is large compared with its radius of curvature, as in [\[link\]](#)(a), we see that the reflected rays do not cross at the same point, and the mirror does not have a well-defined focal point. If the mirror had the shape of a parabola, the rays would all cross at a single point, and the mirror would have a well-defined focal point. But parabolic mirrors are much more expensive to make than spherical mirrors. The solution is to use a mirror that is small compared with its radius of curvature, as shown in [\[link\]](#)(b). (This is the mirror equivalent of the thin lens approximation.) To a very good approximation, this mirror has a well-defined focal point at F that is the focal distance f from the center of the mirror. The focal length f of a concave mirror is positive, since it is a converging mirror.



(a) Parallel rays reflected from a large spherical mirror do not all cross at a common point. (b) If a spherical mirror is small compared with its radius of curvature, parallel rays are focused to a common point. The distance of the focal point from the center of the mirror is its focal length f . Since this mirror is converging, it has a positive focal length.

Just as for lenses, the shorter the focal length, the more powerful the mirror; thus, $P = 1/f$ for a mirror, too. A more strongly curved mirror has a shorter focal length and a greater power. Using the law of reflection and some simple trigonometry, it can be shown that the focal length is half the radius of curvature, or

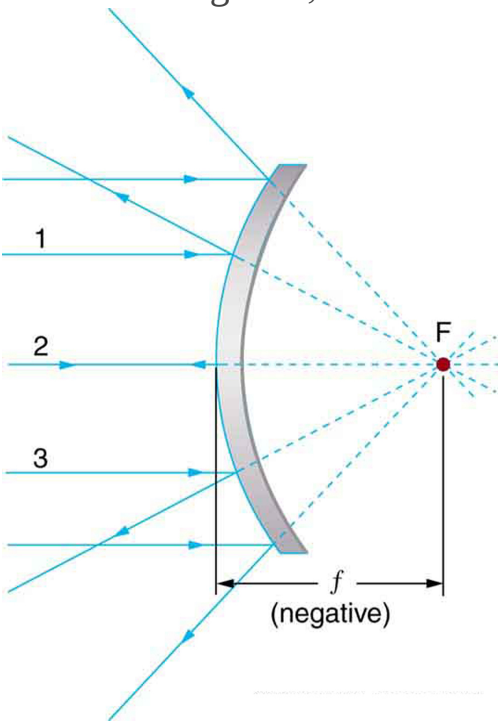
Equation:

$$f = \frac{R}{2},$$

where R is the radius of curvature of a spherical mirror. The smaller the radius of curvature, the smaller the focal length and, thus, the more powerful the mirror.

The convex mirror shown in [\[link\]](#) also has a focal point. Parallel rays of light reflected from the mirror seem to originate from the point F at the

focal distance f behind the mirror. The focal length and power of a convex mirror are negative, since it is a diverging mirror.



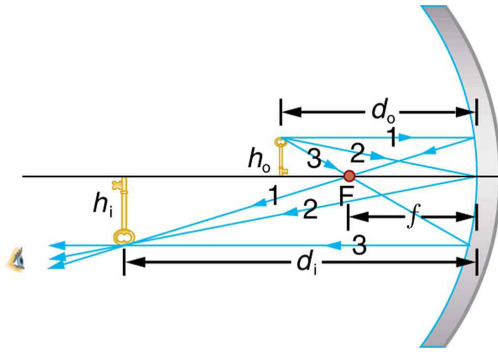
Parallel rays of light reflected from a convex spherical mirror (small in size compared with its radius of curvature) seem to originate from a well-defined focal point at the focal distance f behind the mirror. Convex mirrors diverge light rays and, thus, have a negative focal length.

Ray tracing is as useful for mirrors as for lenses. The rules for ray tracing for mirrors are based on the illustrations just discussed:

1. A ray approaching a concave converging mirror parallel to its axis is reflected through the focal point F of the mirror on the same side. (See rays 1 and 3 in [\[link\]](#)(b).)
2. A ray approaching a convex diverging mirror parallel to its axis is reflected so that it seems to come from the focal point F behind the mirror. (See rays 1 and 3 in [\[link\]](#).)
3. Any ray striking the center of a mirror is followed by applying the law of reflection; it makes the same angle with the axis when leaving as when approaching. (See ray 2 in [\[link\]](#).)
4. A ray approaching a concave converging mirror through its focal point is reflected parallel to its axis. (The reverse of rays 1 and 3 in [\[link\]](#).)
5. A ray approaching a convex diverging mirror by heading toward its focal point on the opposite side is reflected parallel to the axis. (The reverse of rays 1 and 3 in [\[link\]](#).)

We will use ray tracing to illustrate how images are formed by mirrors, and we can use ray tracing quantitatively to obtain numerical information. But since we assume each mirror is small compared with its radius of curvature, we can use the thin lens equations for mirrors just as we did for lenses.

Consider the situation shown in [\[link\]](#), concave spherical mirror reflection, in which an object is placed farther from a concave (converging) mirror than its focal length. That is, f is positive and $d_o > f$, so that we may expect an image similar to the case 1 real image formed by a converging lens. Ray tracing in [\[link\]](#) shows that the rays from a common point on the object all cross at a point on the same side of the mirror as the object. Thus a real image can be projected onto a screen placed at this location. The image distance is positive, and the image is inverted, so its magnification is negative. This is a *case 1 image for mirrors*. It differs from the case 1 image for lenses only in that the image is on the same side of the mirror as the object. It is otherwise identical.



A case 1 image for a mirror. An object is farther from the converging mirror than its focal length. Rays from a common point on the object are traced using the rules in the text. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the mirror, and ray 3 goes through the focal point on the way toward the mirror. All three rays cross at the same point after being reflected, locating the inverted real image. Although three rays are shown, only two of the three are needed to locate the image and determine its height.

Example:

A Concave Reflector

Electric room heaters use a concave mirror to reflect infrared (IR) radiation from hot coils. Note that IR follows the same law of reflection as visible light. Given that the mirror has a radius of curvature of 50.0 cm and produces an image of the coils 3.00 m away from the mirror, where are the coils?

Strategy and Concept

We are given that the concave mirror projects a real image of the coils at an image distance $d_i = 3.00$ m. The coils are the object, and we are asked to find their location—that is, to find the object distance d_o . We are also given the radius of curvature of the mirror, so that its focal length is $f = R/2 = 25.0$ cm (positive since the mirror is concave or converging). Assuming the mirror is small compared with its radius of curvature, we can use the thin lens equations, to solve this problem.

Solution

Since d_i and f are known, thin lens equation can be used to find d_o :

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}.$$

Rearranging to isolate d_o gives

Equation:

$$\frac{1}{d_o} = \frac{1}{f} - \frac{1}{d_i}.$$

Entering known quantities gives a value for $1/d_o$:

Equation:

$$\frac{1}{d_o} = \frac{1}{0.250 \text{ m}} - \frac{1}{3.00 \text{ m}} = \frac{3.667}{\text{m}}.$$

This must be inverted to find d_o :

Equation:

$$d_o = \frac{1 \text{ m}}{3.667} = 27.3 \text{ cm}.$$

Discussion

Note that the object (the filament) is farther from the mirror than the mirror's focal length. This is a case 1 image ($d_o > f$ and f positive), consistent with the fact that a real image is formed. You will get the most concentrated thermal energy directly in front of the mirror and 3.00 m away from it. Generally, this is not desirable, since it could cause burns. Usually, you want the rays to emerge parallel, and this is accomplished by having the filament at the focal point of the mirror.

Note that the filament here is not much farther from the mirror than its focal length and that the image produced is considerably farther away. This is exactly analogous to a slide projector. Placing a slide only slightly farther away from the projector lens than its focal length produces an image significantly farther away. As the object gets closer to the focal distance, the image gets farther away. In fact, as the object distance approaches the focal length, the image distance approaches infinity and the rays are sent out parallel to one another.

Example:

Solar Electric Generating System

One of the solar technologies used today for generating electricity is a device (called a parabolic trough or concentrating collector) that concentrates the sunlight onto a blackened pipe that contains a fluid. This heated fluid is pumped to a heat exchanger, where its heat energy is transferred to another system that is used to generate steam—and so generate electricity through a conventional steam cycle. [\[link\]](#) shows such a working system in southern California. Concave mirrors are used to concentrate the sunlight onto the pipe. The mirror has the approximate shape of a section of a cylinder. For the problem, assume that the mirror is exactly one-quarter of a full cylinder.

- If we wish to place the fluid-carrying pipe 40.0 cm from the concave mirror at the mirror's focal point, what will be the radius of curvature of the mirror?
- Per meter of pipe, what will be the amount of sunlight concentrated onto the pipe, assuming the insolation (incident solar radiation) is

0.900 kW/m²?

- c. If the fluid-carrying pipe has a 2.00-cm diameter, what will be the temperature increase of the fluid per meter of pipe over a period of one minute? Assume all the solar radiation incident on the reflector is absorbed by the pipe, and that the fluid is mineral oil.

Strategy

To solve an *Integrated Concept Problem* we must first identify the physical principles involved. Part (a) is related to the current topic. Part (b) involves a little math, primarily geometry. Part (c) requires an understanding of heat and density.

Solution to (a)

To a good approximation for a concave or semi-spherical surface, the point where the parallel rays from the sun converge will be at the focal point, so $R = 2f = 80.0$ cm.

Solution to (b)

The insolation is 900 W/m^2 . We must find the cross-sectional area A of the concave mirror, since the power delivered is $900 \text{ W/m}^2 \times A$. The mirror in this case is a quarter-section of a cylinder, so the area for a length L of the mirror is $A = \frac{1}{4}(2\pi R)L$. The area for a length of 1.00 m is then

Equation:

$$A = \frac{\pi}{2} R(1.00 \text{ m}) = \frac{(3.14)}{2} (0.800 \text{ m})(1.00 \text{ m}) = 1.26 \text{ m}^2.$$

The insolation on the 1.00-m length of pipe is then

Equation:

$$\left(9.00 \times 10^2 \frac{\text{W}}{\text{m}^2}\right) \left(1.26 \text{ m}^2\right) = 1130 \text{ W}.$$

Solution to (c)

The increase in temperature is given by $Q = mc \Delta T$. The mass m of the mineral oil in the one-meter section of pipe is

Equation:

$$\begin{aligned}
 m &= \rho V = \rho \pi \left(\frac{d}{2} \right)^2 (1.00 \text{ m}) \\
 &= (8.00 \times 10^2 \text{ kg/m}^3)(3.14)(0.0100 \text{ m})^2(1.00 \text{ m}) \\
 &= 0.251 \text{ kg}.
 \end{aligned}$$

Therefore, the increase in temperature in one minute is

Equation:

$$\begin{aligned}
 \Delta T &= Q/mc \\
 &= \frac{(1130 \text{ W})(60.0 \text{ s})}{(0.251 \text{ kg})(1670 \text{ J}\cdot\text{kg}/^\circ\text{C})} \\
 &= 162^\circ\text{C}.
 \end{aligned}$$

Discussion for (c)

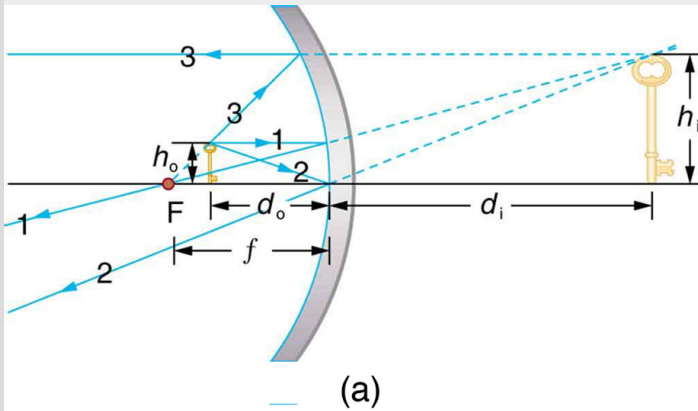
An array of such pipes in the California desert can provide a thermal output of 250 MW on a sunny day, with fluids reaching temperatures as high as 400°C. We are considering only one meter of pipe here, and ignoring heat losses along the pipe.



Parabolic trough collectors are used to generate electricity in southern California. (credit: kjkolb, Wikimedia Commons)

What happens if an object is closer to a concave mirror than its focal length? This is analogous to a case 2 image for lenses ($d_o < f$ and f positive), which is a magnifier. In fact, this is how makeup mirrors act as magnifiers. [\[link\]](#)(a) uses ray tracing to locate the image of an object placed close to a concave mirror. Rays from a common point on the object

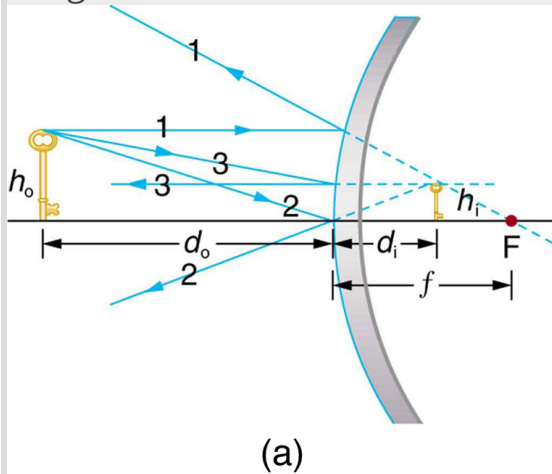
are reflected in such a manner that they appear to be coming from behind the mirror, meaning that the image is virtual and cannot be projected. As with a magnifying glass, the image is upright and larger than the object. This is a *case 2 image for mirrors* and is exactly analogous to that for lenses.



(a) Case 2 images for mirrors are formed when a converging mirror has an object closer to it than its focal length. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the mirror, and ray 3 approaches the mirror as if it came from the focal point. (b) A magnifying mirror showing the reflection. (credit: Mike Melrose, Flickr)

All three rays appear to originate from the same point after being reflected, locating the upright virtual image behind the mirror and showing it to be larger than the object. (b) Makeup mirrors are perhaps the most common use of a concave mirror to produce a larger, upright image.

A convex mirror is a diverging mirror (f is negative) and forms only one type of image. It is a *case 3* image—one that is upright and smaller than the object, just as for diverging lenses. [\[link\]](#)(a) uses ray tracing to illustrate the location and size of the case 3 image for mirrors. Since the image is behind the mirror, it cannot be projected and is thus a virtual image. It is also seen to be smaller than the object.



Case 3 images for mirrors are formed by any convex mirror. Ray 1 approaches parallel to the axis, ray 2 strikes the center of the

mirror, and ray 3 approaches toward the focal point. All three rays appear to originate from the same point after being reflected, locating the upright virtual image behind the mirror and showing it to be smaller than the object.

(b) Security mirrors are convex, producing a smaller, upright image. Because the image is smaller, a larger area is imaged compared to what would be observed for a flat mirror (and hence security is improved).

(credit: Laura D'Alessandro, Flickr)

Example:

Image in a Convex Mirror

A keratometer is a device used to measure the curvature of the cornea, particularly for fitting contact lenses. Light is reflected from the cornea, which acts like a convex mirror, and the keratometer measures the magnification of the image. The smaller the magnification, the smaller the radius of curvature of the cornea. If the light source is 12.0 cm from the cornea and the image's magnification is 0.0320, what is the cornea's radius of curvature?

Strategy

If we can find the focal length of the convex mirror formed by the cornea, we can find its radius of curvature (the radius of curvature is twice the focal length of a spherical mirror). We are given that the object distance is

$d_o = 12.0$ cm and that $m = 0.0320$. We first solve for the image distance d_i , and then for f .

Solution

$m = -d_i/d_o$. Solving this expression for d_i gives

Equation:

$$d_i = -md_o.$$

Entering known values yields

Equation:

$$d_i = -(0.0320)(12.0 \text{ cm}) = -0.384 \text{ cm}.$$

Equation:

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}$$

Substituting known values,

Equation:

$$\frac{1}{f} = \frac{1}{12.0 \text{ cm}} + \frac{1}{-0.384 \text{ cm}} = \frac{-2.52}{\text{cm}}.$$

This must be inverted to find f :

Equation:

$$f = \frac{\text{cm}}{-2.52} = -0.400 \text{ cm}.$$

The radius of curvature is twice the focal length, so that

Equation:

$$R = 2 | f | = 0.800 \text{ cm}.$$

Discussion

Although the focal length f of a convex mirror is defined to be negative, we take the absolute value to give us a positive value for R . The radius of curvature found here is reasonable for a cornea. The distance from cornea

to retina in an adult eye is about 2.0 cm. In practice, many corneas are not spherical, complicating the job of fitting contact lenses. Note that the image distance here is negative, consistent with the fact that the image is behind the mirror, where it cannot be projected. In this section's Problems and Exercises, you will show that for a fixed object distance, the smaller the radius of curvature, the smaller the magnification.

The three types of images formed by mirrors (cases 1, 2, and 3) are exactly analogous to those formed by lenses, as summarized in the table at the end of [Image Formation by Lenses](#). It is easiest to concentrate on only three types of images—then remember that concave mirrors act like convex lenses, whereas convex mirrors act like concave lenses.

Note:**Take-Home Experiment: Concave Mirrors Close to Home**

Find a flashlight and identify the curved mirror used in it. Find another flashlight and shine the first flashlight onto the second one, which is turned off. Estimate the focal length of the mirror. You might try shining a flashlight on the curved mirror behind the headlight of a car, keeping the headlight switched off, and determine its focal length.

Problem-Solving Strategy for Mirrors

Step 1. Examine the situation to determine that image formation by a mirror is involved.

Step 2. Refer to the [Problem-Solving Strategies for Lenses](#). The same strategies are valid for mirrors as for lenses with one qualification—use the ray tracing rules for mirrors listed earlier in this section.

Section Summary

- The characteristics of an image formed by a flat mirror are: (a) The image and object are the same distance from the mirror, (b) The image

- is a virtual image, and (c) The image is situated behind the mirror.
- Image length is half the radius of curvature.

Equation:

$$f = \frac{R}{2}$$

- A convex mirror is a diverging mirror and forms only one type of image, namely a virtual image.

Conceptual Questions

Exercise:

Problem:

What are the differences between real and virtual images? How can you tell (by looking) whether an image formed by a single lens or mirror is real or virtual?

Exercise:

Problem:

Can you see a virtual image? Can you photograph one? Can one be projected onto a screen with additional lenses or mirrors? Explain your responses.

Exercise:

Problem:

Is it necessary to project a real image onto a screen for it to exist?

Exercise:

Problem:

At what distance is an image *always* located—at d_o , d_i , or f ?

Exercise:

Problem:

Under what circumstances will an image be located at the focal point of a lens or mirror?

Exercise:**Problem:**

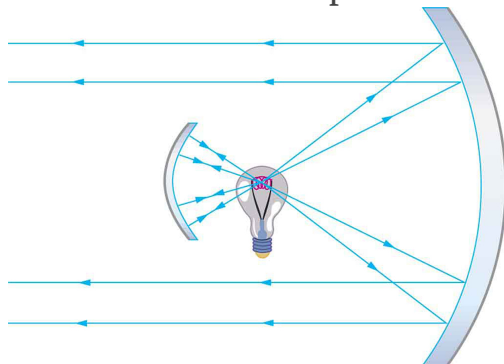
What is meant by a negative magnification? What is meant by a magnification that is less than 1 in magnitude?

Exercise:**Problem:**

Can a case 1 image be larger than the object even though its magnification is always negative? Explain.

Exercise:**Problem:**

[\[link\]](#) shows a light bulb between two mirrors. One mirror produces a beam of light with parallel rays; the other keeps light from escaping without being put into the beam. Where is the filament of the light in relation to the focal point or radius of curvature of each mirror?



The two mirrors trap most of the bulb's light and form a directional beam as in a headlight.

Exercise:**Problem:**

Devise an arrangement of mirrors allowing you to see the back of your head. What is the minimum number of mirrors needed for this task?

Exercise:**Problem:**

If you wish to see your entire body in a flat mirror (from head to toe), how tall should the mirror be? Does its size depend upon your distance away from the mirror? Provide a sketch.

Exercise:**Problem:**

It can be argued that a flat mirror has an infinite focal length. If so, where does it form an image? That is, how are d_i and d_o related?

Exercise:**Problem:**

Why are diverging mirrors often used for rear-view mirrors in vehicles? What is the main disadvantage of using such a mirror compared with a flat one?

Problems & Exercises**Exercise:****Problem:**

What is the focal length of a makeup mirror that has a power of 1.50 D?

Solution:

+0.667 m

Exercise:

Problem:

Some telephoto cameras use a mirror rather than a lens. What radius of curvature mirror is needed to replace a 800 mm focal length telephoto lens?

Exercise:

Problem:

(a) Calculate the focal length of the mirror formed by the shiny back of a spoon that has a 3.00 cm radius of curvature. (b) What is its power in diopters?

Solution:

(a) -1.5×10^{-2} m

(b) -66.7 D

Exercise:

Problem:

Find the magnification of the heater element in [\[link\]](#). Note that its large magnitude helps spread out the reflected energy.

Exercise:

Problem:

What is the focal length of a makeup mirror that produces a magnification of 1.50 when a person's face is 12.0 cm away? Explicitly show how you follow the steps in the [Problem-Solving Strategy for Mirrors](#).

Solution:

+0.360 m (concave)

Exercise:

Problem:

A shopper standing 3.00 m from a convex security mirror sees his image with a magnification of 0.250. (a) Where is his image? (b) What is the focal length of the mirror? (c) What is its radius of curvature? Explicitly show how you follow the steps in the [Problem-Solving Strategy for Mirrors](#).

Exercise:

Problem:

An object 1.50 cm high is held 3.00 cm from a person's cornea, and its reflected image is measured to be 0.167 cm high. (a) What is the magnification? (b) Where is the image? (c) Find the radius of curvature of the convex mirror formed by the cornea. (Note that this technique is used by optometrists to measure the curvature of the cornea for contact lens fitting. The instrument used is called a keratometer, or curve measurer.)

Solution:

(a) +0.111

(b) -0.334 cm (behind “mirror”)

(c) 0.752cm

Exercise:

Problem:

Ray tracing for a flat mirror shows that the image is located a distance behind the mirror equal to the distance of the object from the mirror. This is stated $d_i = -d_o$, since this is a negative image distance (it is a virtual image). (a) What is the focal length of a flat mirror? (b) What is its power?

Exercise:**Problem:**

Show that for a flat mirror $h_i = h_o$, knowing that the image is a distance behind the mirror equal in magnitude to the distance of the object from the mirror.

Solution:**Equation:**

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o} = -\frac{-d_o}{d_o} = \frac{d_o}{d_o} = 1 \Rightarrow h_i = h_o$$

Exercise:**Problem:**

Use the law of reflection to prove that the focal length of a mirror is half its radius of curvature. That is, prove that $f = R/2$. Note this is true for a spherical mirror only if its diameter is small compared with its radius of curvature.

Exercise:**Problem:**

Referring to the electric room heater considered in the first example in this section, calculate the intensity of IR radiation in W/m^2 projected by the concave mirror on a person 3.00 m away. Assume that the heating element radiates 1500 W and has an area of 100 cm^2 , and that half of the radiated power is reflected and focused by the mirror.

Solution:

$$6.82 \text{ kW}/\text{m}^2$$

Exercise:

Problem:

Consider a 250-W heat lamp fixed to the ceiling in a bathroom. If the filament in one light burns out then the remaining three still work. Construct a problem in which you determine the resistance of each filament in order to obtain a certain intensity projected on the bathroom floor. The ceiling is 3.0 m high. The problem will need to involve concave mirrors behind the filaments. Your instructor may wish to guide you on the level of complexity to consider in the electrical components.

Glossary

converging mirror

a concave mirror in which light rays that strike it parallel to its axis converge at one or more points along the axis

diverging mirror

a convex mirror in which light rays that strike it parallel to its axis bend away (diverge) from its axis

law of reflection

angle of reflection equals the angle of incidence

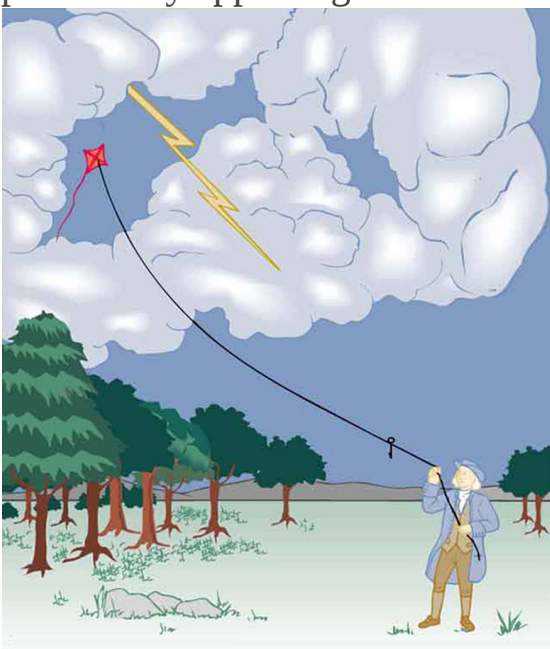
Introduction to Electric Charge and Electric Field

class="introduction"

Static electricity from this plastic slide causes the child's hair to stand on end. The sliding motion stripped electrons away from the child's body, leaving an excess of positive charges, which repel each other along each strand of hair. (credit: Ken Bosma/Wikimedia Commons)



The image of American politician and scientist Benjamin Franklin (1706–1790) flying a kite in a thunderstorm is familiar to every schoolchild. (See [\[link\]](#).) In this experiment, Franklin demonstrated a connection between lightning and **static electricity**. Sparks were drawn from a key hung on a kite string during an electrical storm. These sparks were like those produced by static electricity, such as the spark that jumps from your finger to a metal doorknob after you walk across a wool carpet. What Franklin demonstrated in his dangerous experiment was a connection between phenomena on two different scales: one the grand power of an electrical storm, the other an effect of more human proportions. Connections like this one reveal the underlying unity of the laws of nature, an aspect we humans find particularly appealing.



When Benjamin Franklin demonstrated that lightning was related to static electricity, he made a connection that is now part of the evidence that all directly experienced forces except the gravitational force are manifestations of the electromagnetic force.

Much has been written about Franklin. His experiments were only part of the life of a man who was a scientist, inventor, revolutionary, statesman, and writer. Franklin's experiments were not performed in isolation, nor were they the only ones to reveal connections.

For example, the Italian scientist Luigi Galvani (1737–1798) performed a series of experiments in which static electricity was used to stimulate contractions of leg muscles of dead frogs, an effect already known in humans subjected to static discharges. But Galvani also found that if he joined two metal wires (say copper and zinc) end to end and touched the other ends to muscles, he produced the same effect in frogs as static discharge. Alessandro Volta (1745–1827), partly inspired by Galvani's work, experimented with various combinations of metals and developed the battery.

During the same era, other scientists made progress in discovering fundamental connections. The periodic table was developed as the systematic properties of the elements were discovered. This influenced the development and refinement of the concept of atoms as the basis of matter. Such submicroscopic descriptions of matter also help explain a great deal more.

Atomic and molecular interactions, such as the forces of friction, cohesion, and adhesion, are now known to be manifestations of the **electromagnetic force**. Static electricity is just one aspect of the electromagnetic force, which also includes moving electricity and magnetism.

All the macroscopic forces that we experience directly, such as the sensations of touch and the tension in a rope, are due to the electromagnetic force, one of the four fundamental forces in nature. The gravitational force, another fundamental force, is actually sensed through the electromagnetic interaction of molecules, such as between those in our feet and those on the top of a bathroom scale. (The other two fundamental forces, the strong nuclear force and the weak nuclear force, cannot be sensed on the human scale.)

This chapter begins the study of electromagnetic phenomena at a fundamental level. The next several chapters will cover static electricity, moving electricity, and magnetism—collectively known as electromagnetism. In this chapter, we begin with the study of electric phenomena due to charges that are at least temporarily stationary, called electrostatics, or static electricity.

Glossary

static electricity

a buildup of electric charge on the surface of an object

electromagnetic force

one of the four fundamental forces of nature; the electromagnetic force consists of static electricity, moving electricity and magnetism

Static Electricity and Charge: Conservation of Charge

- Define electric charge, and describe how the two types of charge interact.
- Describe three common situations that generate static electricity.
- State the law of conservation of charge.



Borneo amber was mined in Sabah, Malaysia, from shale-sandstone-mudstone veins.

When a piece of amber is rubbed with a piece of silk, the amber gains more electrons, giving it a net negative charge.

At the same time, the silk, having lost electrons, becomes positively charged. (credit: Sebakoamber, Wikimedia Commons)

What makes plastic wrap cling? Static electricity. Not only are applications of static electricity common these days, its existence has been known since ancient times. The first record of its effects dates to ancient Greeks who noted more than 500 years B.C. that polishing amber temporarily enabled it

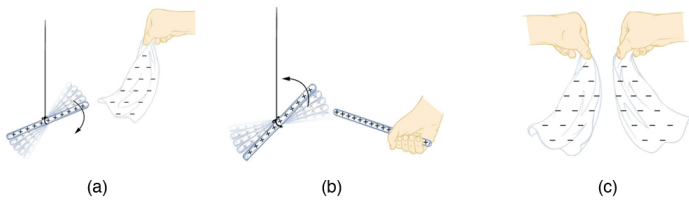
to attract bits of straw (see [\[link\]](#)). The very word *electric* derives from the Greek word for amber (*electron*).

Many of the characteristics of static electricity can be explored by rubbing things together. Rubbing creates the spark you get from walking across a wool carpet, for example. Static cling generated in a clothes dryer and the attraction of straw to recently polished amber also result from rubbing. Similarly, lightning results from air movements under certain weather conditions. You can also rub a balloon on your hair, and the static electricity created can then make the balloon cling to a wall. We also have to be cautious of static electricity, especially in dry climates. When we pump gasoline, we are warned to discharge ourselves (after sliding across the seat) on a metal surface before grabbing the gas nozzle. Attendants in hospital operating rooms must wear booties with aluminum foil on the bottoms to avoid creating sparks which may ignite the oxygen being used.

Some of the most basic characteristics of static electricity include:

- The effects of static electricity are explained by a physical quantity not previously introduced, called electric charge.
- There are only two types of charge, one called positive and the other called negative.
- Like charges repel, whereas unlike charges attract.
- The force between charges decreases with distance.

How do we know there are two types of **electric charge**? When various materials are rubbed together in controlled ways, certain combinations of materials always produce one type of charge on one material and the opposite type on the other. By convention, we call one type of charge “positive”, and the other type “negative.” For example, when glass is rubbed with silk, the glass becomes positively charged and the silk negatively charged. Since the glass and silk have opposite charges, they attract one another like clothes that have rubbed together in a dryer. Two glass rods rubbed with silk in this manner will repel one another, since each rod has positive charge on it. Similarly, two silk cloths so rubbed will repel, since both cloths have negative charge. [\[link\]](#) shows how these simple materials can be used to explore the nature of the force between charges.



A glass rod becomes positively charged when rubbed with silk, while the silk becomes negatively charged.

(a) The glass rod is attracted to the silk because their charges are opposite. (b) Two similarly charged glass rods repel. (c) Two similarly charged silk cloths repel.

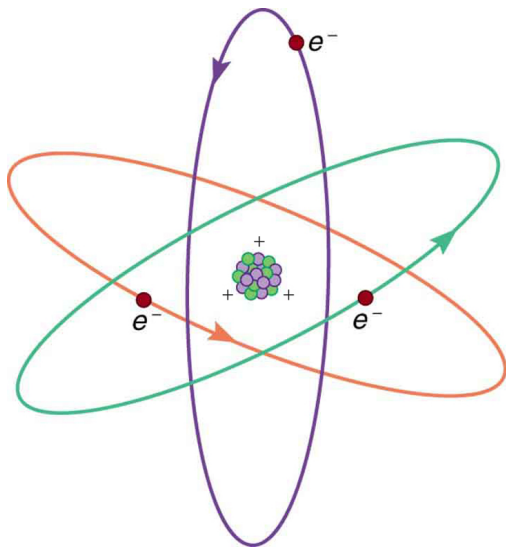
More sophisticated questions arise. Where do these charges come from? Can you create or destroy charge? Is there a smallest unit of charge? Exactly how does the force depend on the amount of charge and the distance between charges? Such questions obviously occurred to Benjamin Franklin and other early researchers, and they interest us even today.

Charge Carried by Electrons and Protons

Franklin wrote in his letters and books that he could see the effects of electric charge but did not understand what caused the phenomenon. Today we have the advantage of knowing that normal matter is made of atoms, and that atoms contain positive and negative charges, usually in equal amounts.

[\[link\]](#) shows a simple model of an atom with negative **electrons** orbiting its positive nucleus. The nucleus is positive due to the presence of positively charged **protons**. Nearly all charge in nature is due to electrons and protons, which are two of the three building blocks of most matter. (The third is the neutron, which is neutral, carrying no charge.) Other charge-carrying particles are observed in cosmic rays and nuclear decay, and are created in

particle accelerators. All but the electron and proton survive only a short time and are quite rare by comparison.



This simplified (and not to scale) view of an atom is called the planetary model of the atom.

Negative electrons orbit a much heavier positive nucleus, as the planets orbit the much heavier sun. There the similarity ends, because forces in the atom are electromagnetic, whereas those in the planetary system are gravitational.

Normal macroscopic amounts of matter contain immense numbers of atoms and molecules and, hence, even greater numbers of individual

negative and positive charges.

The charges of electrons and protons are identical in magnitude but opposite in sign. Furthermore, all charged objects in nature are integral multiples of this basic quantity of charge, meaning that all charges are made of combinations of a basic unit of charge. Usually, charges are formed by combinations of electrons and protons. The magnitude of this basic charge is

Equation:

$$|q_e| = 1.60 \times 10^{-19} \text{ C}.$$

The symbol q is commonly used for charge and the subscript e indicates the charge of a single electron (or proton).

The SI unit of charge is the coulomb (C). The number of protons needed to make a charge of 1.00 C is

Equation:

$$1.00 \text{ C} \times \frac{1 \text{ proton}}{1.60 \times 10^{-19} \text{ C}} = 6.25 \times 10^{18} \text{ protons}.$$

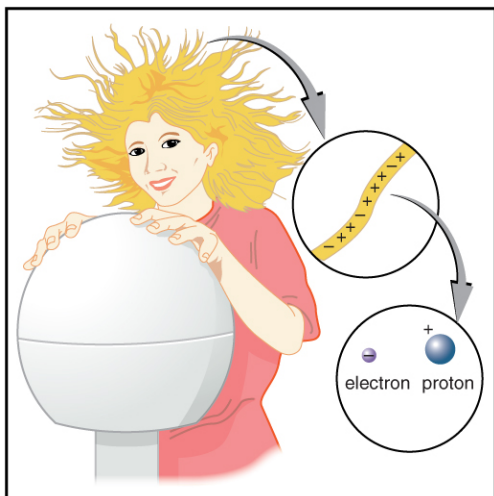
Similarly, 6.25×10^{18} electrons have a combined charge of -1.00 coulomb. Just as there is a smallest bit of an element (an atom), there is a smallest bit of charge. There is no directly observed charge smaller than $|q_e|$ (see [Things Great and Small: The Submicroscopic Origin of Charge](#)), and all observed charges are integral multiples of $|q_e|$.

Note:

Things Great and Small: The Submicroscopic Origin of Charge

With the exception of exotic, short-lived particles, all charge in nature is carried by electrons and protons. Electrons carry the charge we have named negative. Protons carry an equal-magnitude charge that we call positive. (See [\[link\]](#).) Electron and proton charges are considered fundamental building blocks, since all other charges are integral multiples of those carried by electrons and protons. Electrons and protons are also two of the three fundamental building blocks of ordinary matter. The neutron is the third and has zero total charge.

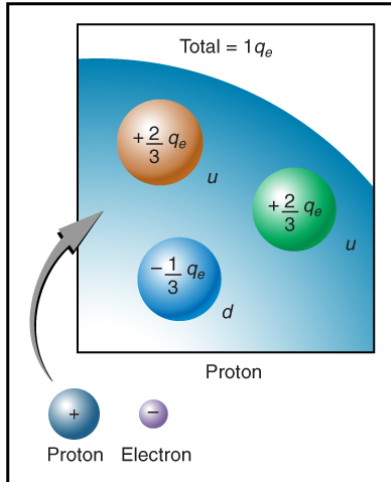
[\[link\]](#) shows a person touching a Van de Graaff generator and receiving excess positive charge. The expanded view of a hair shows the existence of both types of charges but an excess of positive. The repulsion of these positive like charges causes the strands of hair to repel other strands of hair and to stand up. The further blowup shows an artist's conception of an electron and a proton perhaps found in an atom in a strand of hair.



When this person touches a Van de Graaff generator, she receives an excess of positive charge, causing her hair to stand on end. The charges in

one hair are shown. An artist's conception of an electron and a proton illustrate the particles carrying the negative and positive charges. We cannot really see these particles with visible light because they are so small (the electron seems to be an infinitesimal point), but we know a great deal about their measurable properties, such as the charges they carry.

The electron seems to have no substructure; in contrast, when the substructure of protons is explored by scattering extremely energetic electrons from them, it appears that there are point-like particles inside the proton. These sub-particles, named quarks, have never been directly observed, but they are believed to carry fractional charges as seen in [\[link\]](#). Charges on electrons and protons and all other directly observable particles are unitary, but these quark substructures carry charges of either $-\frac{1}{3}$ or $+\frac{2}{3}$. There are continuing attempts to observe fractional charge directly and to learn of the properties of quarks, which are perhaps the ultimate substructure of matter.



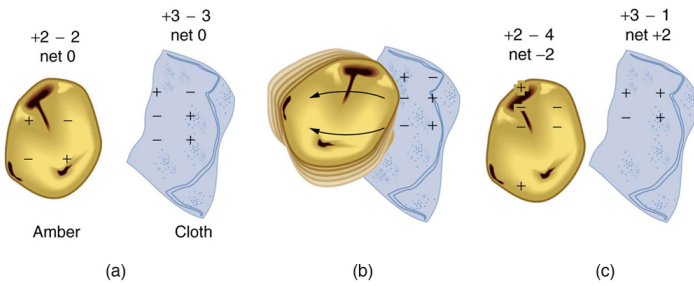
Artist's conception of fractional quark charges inside a proton. A group of three quark charges add up to the single positive charge on the proton:

$$-\frac{1}{3}q_e + \frac{2}{3}q_e + \frac{2}{3}q_e = +1q_e$$

.

Separation of Charge in Atoms

Charges in atoms and molecules can be separated—for example, by rubbing materials together. Some atoms and molecules have a greater affinity for electrons than others and will become negatively charged by close contact in rubbing, leaving the other material positively charged. (See [link](#).) Positive charge can similarly be induced by rubbing. Methods other than rubbing can also separate charges. Batteries, for example, use combinations of substances that interact in such a way as to separate charges. Chemical interactions may transfer negative charge from one substance to the other, making one battery terminal negative and leaving the first one positive.



When materials are rubbed together, charges can be separated, particularly if one material has a greater affinity for electrons than another. (a) Both the amber and cloth are originally neutral, with equal positive and negative charges. Only a tiny fraction of the charges are involved, and only a few of them are shown here. (b) When rubbed together, some negative charge is transferred to the amber, leaving the cloth with a net positive charge. (c) When separated, the amber and cloth now have net charges, but the absolute value of the net positive and negative charges will be equal.

No charge is actually created or destroyed when charges are separated as we have been discussing. Rather, existing charges are moved about. In fact, in all situations the total amount of charge is always constant. This universally obeyed law of nature is called the **law of conservation of charge**.

Note:

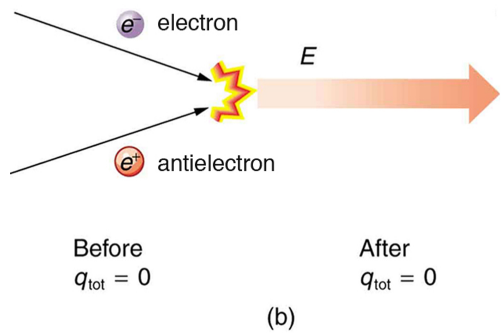
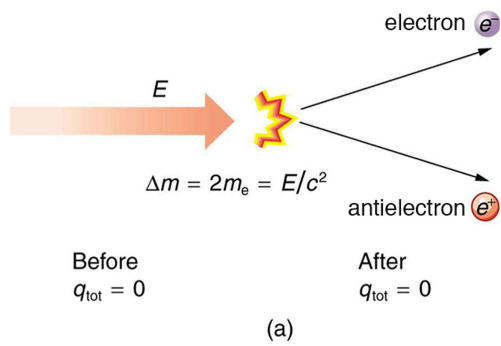
Law of Conservation of Charge

Total charge is constant in any process.

In more exotic situations, such as in particle accelerators, mass, Δm , can be created from energy in the amount $\Delta m = \frac{E}{c^2}$. Sometimes, the created mass is charged, such as when an electron is created. Whenever a charged particle is created, another having an opposite charge is always created along with it, so that the total charge created is zero. Usually, the two particles are “matter-antimatter” counterparts. For example, an antielectron would usually be created at the same time as an electron. The antielectron has a positive charge (it is called a positron), and so the total charge created is zero. (See [\[link\]](#).) All particles have antimatter counterparts with opposite signs. When matter and antimatter counterparts are brought together, they completely annihilate one another. By annihilate, we mean that the mass of the two particles is converted to energy E , again obeying the relationship $\Delta m = \frac{E}{c^2}$. Since the two particles have equal and opposite charge, the total charge is zero before and after the annihilation; thus, total charge is conserved.

Note:**Making Connections: Conservation Laws**

Only a limited number of physical quantities are universally conserved. Charge is one—energy, momentum, and angular momentum are others. Because they are conserved, these physical quantities are used to explain more phenomena and form more connections than other, less basic quantities. We find that conserved quantities give us great insight into the rules followed by nature and hints to the organization of nature. Discoveries of conservation laws have led to further discoveries, such as the weak nuclear force and the quark substructure of protons and other particles.



(a) When enough energy is present, it can be converted into matter. Here the matter created is an electron–antielectron pair. (m_e is the electron’s mass.) The total charge before and after this event is zero. (b) When matter and antimatter collide, they annihilate each other; the total charge is conserved at zero before and after the annihilation.

The law of conservation of charge is absolute—it has never been observed to be violated. Charge, then, is a special physical quantity, joining a very

short list of other quantities in nature that are always conserved. Other conserved quantities include energy, momentum, and angular momentum.

Note:

PhET Explorations: Balloons and Static Electricity

Why does a balloon stick to your sweater? Rub a balloon on a sweater, then let go of the balloon and it flies over and sticks to the sweater. View the charges in the sweater, balloons, and the wall.

https://phet.colorado.edu/sims/html/balloons-and-static-electricity/latest/balloons-and-static-electricity_en.html

Section Summary

- There are only two types of charge, which we call positive and negative.
- Like charges repel, unlike charges attract, and the force between charges decreases with the square of the distance.
- The vast majority of positive charge in nature is carried by protons, while the vast majority of negative charge is carried by electrons.
- The electric charge of one electron is equal in magnitude and opposite in sign to the charge of one proton.
- An ion is an atom or molecule that has nonzero total charge due to having unequal numbers of electrons and protons.
- The SI unit for charge is the coulomb (C), with protons and electrons having charges of opposite sign but equal magnitude; the magnitude of this basic charge $|q_e|$ is

Equation:

$$|q_e| = 1.60 \times 10^{-19} \text{ C}.$$

- Whenever charge is created or destroyed, equal amounts of positive and negative are involved.
- Most often, existing charges are separated from neutral objects to obtain some net charge.

- Both positive and negative charges exist in neutral objects and can be separated by rubbing one object with another. For macroscopic objects, negatively charged means an excess of electrons and positively charged means a depletion of electrons.
- The law of conservation of charge ensures that whenever a charge is created, an equal charge of the opposite sign is created at the same time.

Conceptual Questions

Exercise:

Problem:

There are very large numbers of charged particles in most objects. Why, then, don't most objects exhibit static electricity?

Exercise:

Problem:

Why do most objects tend to contain nearly equal numbers of positive and negative charges?

Problems & Exercises

Exercise:

Problem:

Common static electricity involves charges ranging from nanocoulombs to microcoulombs. (a) How many electrons are needed to form a charge of -2.00 nC (b) How many electrons must be removed from a neutral object to leave a net charge of $0.500 \mu\text{C}$?

Solution:

(a) 1.25×10^{10}

(b) 3.13×10^{12}

Exercise:

Problem:

If 1.80×10^{20} electrons move through a pocket calculator during a full day's operation, how many coulombs of charge moved through it?

Exercise:

Problem:

To start a car engine, the car battery moves 3.75×10^{21} electrons through the starter motor. How many coulombs of charge were moved?

Solution:

-600 C

Exercise:

Problem:

A certain lightning bolt moves 40.0 C of charge. How many fundamental units of charge $|q_e|$ is this?

Glossary

electric charge

a physical property of an object that causes it to be attracted toward or repelled from another charged object; each charged object generates and is influenced by a force called an electromagnetic force

law of conservation of charge

states that whenever a charge is created, an equal amount of charge with the opposite sign is created simultaneously

electron

a particle orbiting the nucleus of an atom and carrying the smallest unit of negative charge

proton

a particle in the nucleus of an atom and carrying a positive charge equal in magnitude and opposite in sign to the amount of negative charge carried by an electron

Conductors and Insulators

- Define conductor and insulator, explain the difference, and give examples of each.
- Describe three methods for charging an object.
- Explain what happens to an electric force as you move farther from the source.
- Define polarization.

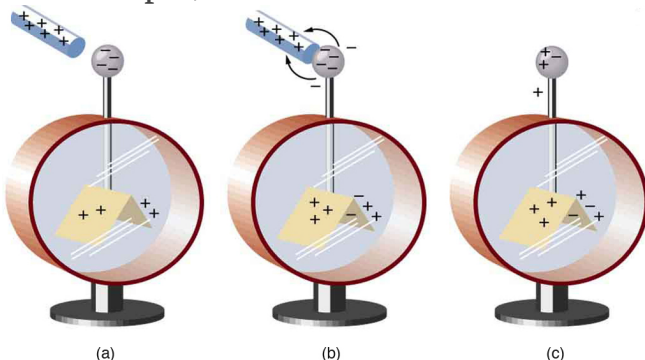


This power adapter uses metal wires and connectors to conduct electricity from the wall socket to a laptop computer. The conducting wires allow electrons to move freely through the cables, which are shielded by rubber and plastic. These materials act as insulators that don't allow electric charge to escape outward. (credit: Evan-Amos, Wikimedia Commons)

Some substances, such as metals and salty water, allow charges to move through them with relative ease. Some of the electrons in metals and similar conductors are not bound to individual atoms or sites in the material. These **free electrons** can move through the material much as air moves through loose sand. Any substance that has free electrons and allows charge to move

relatively freely through it is called a **conductor**. The moving electrons may collide with fixed atoms and molecules, losing some energy, but they can move in a conductor. Superconductors allow the movement of charge without any loss of energy. Salty water and other similar conducting materials contain free ions that can move through them. An ion is an atom or molecule having a positive or negative (nonzero) total charge. In other words, the total number of electrons is not equal to the total number of protons.

Other substances, such as glass, do not allow charges to move through them. These are called **insulators**. Electrons and ions in insulators are bound in the structure and cannot move easily—as much as 10^{23} times more slowly than in conductors. Pure water and dry table salt are insulators, for example, whereas molten salt and salty water are conductors.



An electroscope is a favorite instrument in physics demonstrations and student laboratories. It is typically made with gold foil leaves hung from a (conducting) metal stem and is insulated from the room air in a glass-walled container. (a) A positively charged glass rod is brought near the tip of the electroscope, attracting electrons to the top and leaving a net positive charge on the leaves. Like charges in the light flexible gold leaves

repel, separating them. (b) When the rod is touched against the ball, electrons are attracted and transferred, reducing the net charge on the glass rod but leaving the electroscope positively charged. (c) The excess charges are evenly distributed in the stem and leaves of the electroscope once the glass rod is removed.

Charging by Contact

[\[link\]](#) shows an electroscope being charged by touching it with a positively charged glass rod. Because the glass rod is an insulator, it must actually touch the electroscope to transfer charge to or from it. (Note that the extra positive charges reside on the surface of the glass rod as a result of rubbing it with silk before starting the experiment.) Since only electrons move in metals, we see that they are attracted to the top of the electroscope. There, some are transferred to the positive rod by touch, leaving the electroscope with a net positive charge.

Electrostatic repulsion in the leaves of the charged electroscope separates them. The electrostatic force has a horizontal component that results in the leaves moving apart as well as a vertical component that is balanced by the gravitational force. Similarly, the electroscope can be negatively charged by contact with a negatively charged object.

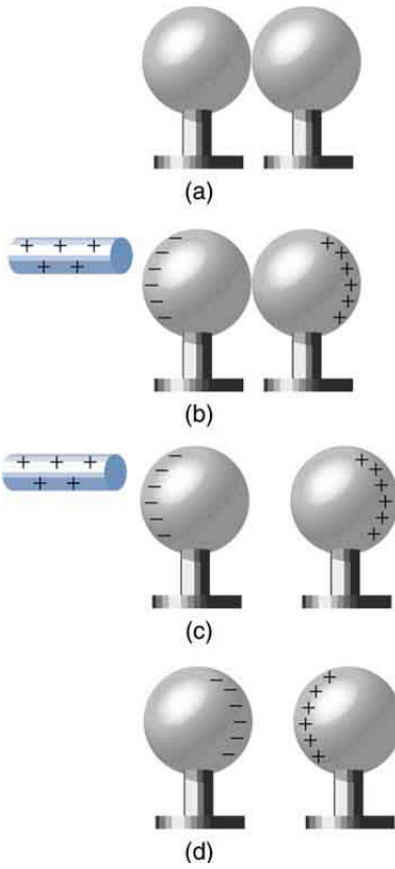
Charging by Induction

It is not necessary to transfer excess charge directly to an object in order to charge it. [\[link\]](#) shows a method of **induction** wherein a charge is created in a nearby object, without direct contact. Here we see two neutral metal spheres in contact with one another but insulated from the rest of the world.

A positively charged rod is brought near one of them, attracting negative charge to that side, leaving the other sphere positively charged.

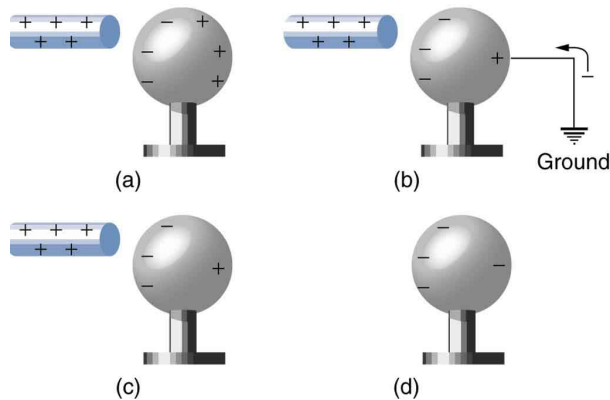
This is an example of induced **polarization** of neutral objects. Polarization is the separation of charges in an object that remains neutral. If the spheres are now separated (before the rod is pulled away), each sphere will have a net charge. Note that the object closest to the charged rod receives an opposite charge when charged by induction. Note also that no charge is removed from the charged rod, so that this process can be repeated without depleting the supply of excess charge.

Another method of charging by induction is shown in [\[link\]](#). The neutral metal sphere is polarized when a charged rod is brought near it. The sphere is then grounded, meaning that a conducting wire is run from the sphere to the ground. Since the earth is large and most ground is a good conductor, it can supply or accept excess charge easily. In this case, electrons are attracted to the sphere through a wire called the ground wire, because it supplies a conducting path to the ground. The ground connection is broken before the charged rod is removed, leaving the sphere with an excess charge opposite to that of the rod. Again, an opposite charge is achieved when charging by induction and the charged rod loses none of its excess charge.



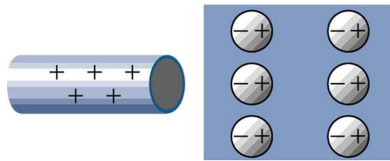
Charging by induction. (a) Two uncharged or neutral metal spheres are in contact with each other but insulated from the rest of the world. (b) A positively charged glass rod is brought near the sphere on the left, attracting negative charge and leaving the other sphere positively charged. (c) The

spheres are separated before the rod is removed, thus separating negative and positive charge. (d) The spheres retain net charges after the inducing rod is removed—without ever having been touched by a charged object.

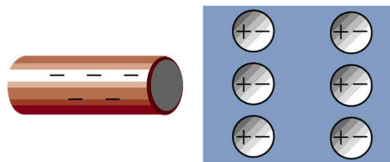


Charging by induction, using a ground connection. (a) A positively charged rod is brought near a neutral metal sphere, polarizing it. (b) The sphere is grounded, allowing electrons to be attracted from the earth's ample supply. (c) The ground connection is broken. (d) The positive rod is

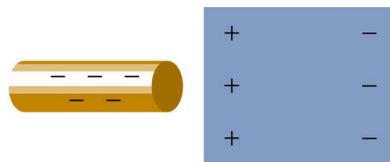
removed, leaving the sphere
with an induced negative
charge.



(a)



(b)



(c)

Both positive and
negative objects
attract a neutral
object by polarizing
its molecules. (a) A
positive object
brought near a
neutral insulator
polarizes its
molecules. There is
a slight shift in the
distribution of the
electrons orbiting
the molecule, with

unlike charges being brought nearer and like charges moved away. Since the electrostatic force decreases with distance, there is a net attraction. (b) A negative object produces the opposite polarization, but again attracts the neutral object. (c) The same effect occurs for a conductor; since the unlike charges are closer, there is a net attraction.

Neutral objects can be attracted to any charged object. The pieces of straw attracted to polished amber are neutral, for example. If you run a plastic comb through your hair, the charged comb can pick up neutral pieces of paper. [\[link\]](#) shows how the polarization of atoms and molecules in neutral objects results in their attraction to a charged object.

When a charged rod is brought near a neutral substance, an insulator in this case, the distribution of charge in atoms and molecules is shifted slightly. Opposite charge is attracted nearer the external charged rod, while like charge is repelled. Since the electrostatic force decreases with distance, the repulsion of like charges is weaker than the attraction of unlike charges, and so there is a net attraction. Thus a positively charged glass rod attracts neutral pieces of paper, as will a negatively charged rubber rod. Some

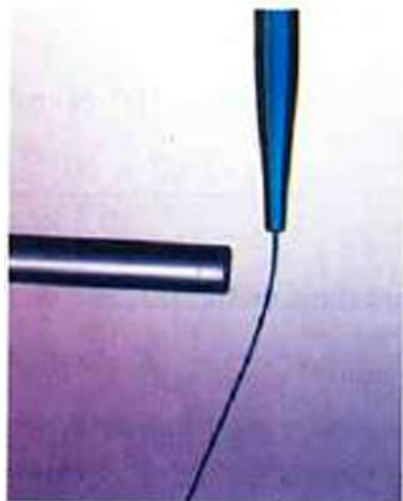
molecules, like water, are polar molecules. Polar molecules have a natural or inherent separation of charge, although they are neutral overall. Polar molecules are particularly affected by other charged objects and show greater polarization effects than molecules with naturally uniform charge distributions.

Exercise:

Check Your Understanding

Problem:

Can you explain the attraction of water to the charged rod in the figure below?



Solution:

Answer

Water molecules are polarized, giving them slightly positive and slightly negative sides. This makes water even more susceptible to a charged rod's attraction. As the water flows downward, due to the force of gravity, the charged conductor exerts a net attraction to the opposite charges in the stream of water, pulling it closer.

Note:

PhET Explorations: John Travoltage

Make sparks fly with John Travoltage. Wiggle Johnnie's foot and he picks up charges from the carpet. Bring his hand close to the door knob and get rid of the excess charge.

https://phet.colorado.edu/sims/html/john-travoltage/latest/john-travoltage_en.html

Section Summary

- Polarization is the separation of positive and negative charges in a neutral object.
- A conductor is a substance that allows charge to flow freely through its atomic structure.
- An insulator holds charge within its atomic structure.
- Objects with like charges repel each other, while those with unlike charges attract each other.
- A conducting object is said to be grounded if it is connected to the Earth through a conductor. Grounding allows transfer of charge to and from the earth's large reservoir.
- Objects can be charged by contact with another charged object and obtain the same sign charge.
- If an object is temporarily grounded, it can be charged by induction, and obtains the opposite sign charge.
- Polarized objects have their positive and negative charges concentrated in different areas, giving them a non-symmetrical charge.
- Polar molecules have an inherent separation of charge.

Conceptual Questions

Exercise:

Problem:

An eccentric inventor attempts to levitate by first placing a large negative charge on himself and then putting a large positive charge on the ceiling of his workshop. Instead, while attempting to place a large negative charge on himself, his clothes fly off. Explain.

Exercise:**Problem:**

If you have charged an electroscope by contact with a positively charged object, describe how you could use it to determine the charge of other objects. Specifically, what would the leaves of the electroscope do if other charged objects were brought near its knob?

Exercise:**Problem:**

When a glass rod is rubbed with silk, it becomes positive and the silk becomes negative—yet both attract dust. Does the dust have a third type of charge that is attracted to both positive and negative? Explain.

Exercise:**Problem:**

Why does a car always attract dust right after it is polished? (Note that car wax and car tires are insulators.)

Exercise:**Problem:**

Describe how a positively charged object can be used to give another object a negative charge. What is the name of this process?

Exercise:**Problem:**

What is grounding? What effect does it have on a charged conductor? On a charged insulator?

Problems & Exercises**Exercise:**

Problem:

Suppose a speck of dust in an electrostatic precipitator has 1.0000×10^{12} protons in it and has a net charge of -5.00 nC (a very large charge for a small speck). How many electrons does it have?

Solution:

$$1.03 \times 10^{12}$$

Exercise:**Problem:**

An amoeba has 1.00×10^{16} protons and a net charge of 0.300 pC . (a) How many fewer electrons are there than protons? (b) If you paired them up, what fraction of the protons would have no electrons?

Exercise:**Problem:**

A 50.0 g ball of copper has a net charge of $2.00 \mu\text{C}$. What fraction of the copper's electrons has been removed? (Each copper atom has 29 protons, and copper has an atomic mass of 63.5.)

Solution:

$$9.09 \times 10^{-13}$$

Exercise:**Problem:**

What net charge would you place on a 100 g piece of sulfur if you put an extra electron on $1 \text{ in } 10^{12}$ of its atoms? (Sulfur has an atomic mass of 32.1.)

Exercise:

Problem:

How many coulombs of positive charge are there in 4.00 kg of plutonium, given its atomic mass is 244 and that each plutonium atom has 94 protons?

Solution:

$$1.48 \times 10^8 \text{ C}$$

Glossary

free electron

an electron that is free to move away from its atomic orbit

conductor

a material that allows electrons to move separately from their atomic orbits

insulator

a material that holds electrons securely within their atomic orbits

grounded

when a conductor is connected to the Earth, allowing charge to freely flow to and from Earth's unlimited reservoir

induction

the process by which an electrically charged object brought near a neutral object creates a charge in that object

polarization

slight shifting of positive and negative charges to opposite sides of an atom or molecule

electrostatic repulsion

the phenomenon of two objects with like charges repelling each other

Coulomb's Law

- State Coulomb's law in terms of how the electrostatic force changes with the distance between two objects.
- Calculate the electrostatic force between two charged point forces, such as electrons or protons.
- Compare the electrostatic force to the gravitational attraction for a proton and an electron; for a human and the Earth.



This NASA image of Arp 87 shows the result of a strong gravitational attraction between two galaxies. In contrast, at the subatomic level, the electrostatic attraction between two objects, such as an electron and a proton, is far greater than their mutual attraction due to gravity. (credit: NASA/HST)

Through the work of scientists in the late 18th century, the main features of the **electrostatic force**—the existence of two types of charge, the observation that like charges repel, unlike charges attract, and the decrease of force with distance—were eventually refined, and expressed as a mathematical formula. The mathematical formula for the electrostatic force is called **Coulomb's law** after the French physicist Charles Coulomb (1736–1806), who performed experiments and first proposed a formula to calculate it.

Note:

Coulomb's Law

Equation:

$$F = k \frac{|q_1 q_2|}{r^2}.$$

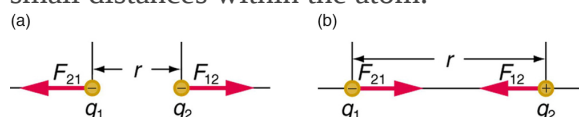
Coulomb's law calculates the magnitude of the force F between two point charges, q_1 and q_2 , separated by a distance r . In SI units, the constant k is equal to

Equation:

$$k = 8.988 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2} \approx 8.99 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2}.$$

The electrostatic force is a vector quantity and is expressed in units of newtons. The force is understood to be along the line joining the two charges. (See [\[link\]](#).)

Although the formula for Coulomb's law is simple, it was no mean task to prove it. The experiments Coulomb did, with the primitive equipment then available, were difficult. Modern experiments have verified Coulomb's law to great precision. For example, it has been shown that the force is inversely proportional to distance between two objects squared ($F \propto 1/r^2$) to an accuracy of 1 part in 10^{16} . No exceptions have ever been found, even at the small distances within the atom.



The magnitude of the electrostatic force F between point charges q_1 and q_2 separated by a distance r is given

by Coulomb's law. Note that

Newton's third law (every force exerted creates an equal and opposite force) applies as usual—the force on q_1 is equal in magnitude and opposite in direction to the force it exerts on q_2 .

(a) Like charges. (b) Unlike charges.

Example:

How Strong is the Coulomb Force Relative to the Gravitational Force?

Compare the electrostatic force between an electron and proton separated by 0.530×10^{-10} m with the gravitational force between them. This distance is their average separation in a hydrogen atom.

Strategy

To compare the two forces, we first compute the electrostatic force using Coulomb's law, $F = k \frac{|q_1 q_2|}{r^2}$. We then calculate the gravitational force using Newton's universal law of

gravitation. Finally, we take a ratio to see how the forces compare in magnitude.

Solution

Entering the given and known information about the charges and separation of the electron and proton into the expression of Coulomb's law yields

Equation:

$$F = k \frac{|q_1 q_2|}{r^2}$$

Equation:

$$= (8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \times \frac{(1.60 \times 10^{-19} \text{ C})(1.60 \times 10^{-19} \text{ C})}{(0.530 \times 10^{-10} \text{ m})^2}$$

Thus the Coulomb force is

Equation:

$$F = 8.19 \times 10^{-8} \text{ N}.$$

The charges are opposite in sign, so this is an attractive force. This is a very large force for an electron—it would cause an acceleration of $8.99 \times 10^{22} \text{ m/s}^2$ (verification is left as an end-of-section problem). The gravitational force is given by Newton's law of gravitation as:

Equation:

$$F_G = G \frac{mM}{r^2},$$

where $G = 6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$. Here m and M represent the electron and proton masses, which can be found in the appendices. Entering values for the knowns yields

Equation:

$$F_G = (6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2) \times \frac{(9.11 \times 10^{-31} \text{ kg})(1.67 \times 10^{-27} \text{ kg})}{(0.530 \times 10^{-10} \text{ m})^2} = 3.61 \times 10^{-47} \text{ N}$$

This is also an attractive force, although it is traditionally shown as positive since gravitational force is always attractive. The ratio of the magnitude of the electrostatic force to gravitational force in this case is, thus,

Equation:

$$\frac{F}{F_G} = 2.27 \times 10^{39}.$$

Discussion

This is a remarkably large ratio! Note that this will be the ratio of electrostatic force to gravitational force for an electron and a proton at any distance (taking the ratio before entering numerical values shows that the distance cancels). This ratio gives some indication

of just how much larger the Coulomb force is than the gravitational force between two of the most common particles in nature.

As the example implies, gravitational force is completely negligible on a small scale, where the interactions of individual charged particles are important. On a large scale, such as between the Earth and a person, the reverse is true. Most objects are nearly electrically neutral, and so attractive and repulsive **Coulomb forces** nearly cancel. Gravitational force on a large scale dominates interactions between large objects because it is always attractive, while Coulomb forces tend to cancel.

Section Summary

- Frenchman Charles Coulomb was the first to publish the mathematical equation that describes the electrostatic force between two objects.
- Coulomb's law gives the magnitude of the force between point charges. It is **Equation:**

$$F = k \frac{|q_1 q_2|}{r^2},$$

where q_1 and q_2 are two point charges separated by a distance r , and $k \approx 8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$

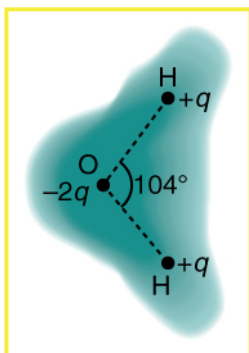
- This Coulomb force is extremely basic, since most charges are due to point-like particles. It is responsible for all electrostatic effects and underlies most macroscopic forces.
- The Coulomb force is extraordinarily strong compared with the gravitational force, another basic force—but unlike gravitational force it can cancel, since it can be either attractive or repulsive.
- The electrostatic force between two subatomic particles is far greater than the gravitational force between the same two particles.

Conceptual Questions

Exercise:

Problem:

[\[link\]](#) shows the charge distribution in a water molecule, which is called a polar molecule because it has an inherent separation of charge. Given water's polar character, explain what effect humidity has on removing excess charge from objects.



Schematic representation of the outer electron cloud of a neutral water molecule. The electrons spend more time near the oxygen than the hydrogens, giving a permanent charge separation as shown. Water is thus a *polar molecule*. It is more easily affected by electrostatic forces than molecules with uniform charge distributions.

Exercise:**Problem:**

Using [\[link\]](#), explain, in terms of Coulomb's law, why a polar molecule (such as in [\[link\]](#)) is attracted by both positive and negative charges.

Exercise:

Problem:

Given the polar character of water molecules, explain how ions in the air form nucleation centers for rain droplets.

Problems & Exercises

Exercise:

Problem:

What is the repulsive force between two pith balls that are 8.00 cm apart and have equal charges of -30.0 nC ?

Exercise:

Problem:

(a) How strong is the attractive force between a glass rod with a $0.700\text{ }\mu\text{C}$ charge and a silk cloth with a $-0.600\text{ }\mu\text{C}$ charge, which are 12.0 cm apart, using the approximation that they act like point charges? (b) Discuss how the answer to this problem might be affected if the charges are distributed over some area and do not act like point charges.

Solution:

(a) 0.263 N

(b) If the charges are distributed over some area, there will be a concentration of charge along the side closest to the oppositely charged object. This effect will increase the net force.

Exercise:

Problem:

Two point charges exert a 5.00 N force on each other. What will the force become if the distance between them is increased by a factor of three?

Exercise:

Problem:

Two point charges are brought closer together, increasing the force between them by a factor of 25. By what factor was their separation decreased?

Solution:

The separation decreased by a factor of 5.

Exercise:**Problem:**

How far apart must two point charges of 75.0 nC (typical of static electricity) be to have a force of 1.00 N between them?

Exercise:**Problem:**

If two equal charges each of 1 C each are separated in air by a distance of 1 km, what is the magnitude of the force acting between them? You will see that even at a distance as large as 1 km, the repulsive force is substantial because 1 C is a very significant amount of charge.

Exercise:**Problem:**

A test charge of $+2\ \mu\text{C}$ is placed halfway between a charge of $+6\ \mu\text{C}$ and another of $+4\ \mu\text{C}$ separated by 10 cm. (a) What is the magnitude of the force on the test charge? (b) What is the direction of this force (away from or toward the $+6\ \mu\text{C}$ charge)?

Exercise:**Problem:**

Bare free charges do not remain stationary when close together. To illustrate this, calculate the acceleration of two isolated protons separated by 2.00 nm (a typical distance between gas atoms). Explicitly show how you follow the steps in the Problem-Solving Strategy for electrostatics.

Solution:

$$\begin{aligned} F &= k \frac{|q_1 q_2|}{r^2} = ma \Rightarrow a = \frac{kq^2}{mr^2} \\ &= \frac{(9.00 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2)(1.60 \times 10^{-19} \text{ m})^2}{(1.67 \times 10^{-27} \text{ kg})(2.00 \times 10^{-9} \text{ m})^2} \\ &= 3.45 \times 10^{16} \text{ m/s}^2 \end{aligned}$$

Exercise:**Problem:**

(a) By what factor must you change the distance between two point charges to change the force between them by a factor of 10? (b) Explain how the distance can either increase or decrease by this factor and still cause a factor of 10 change in the force.

Solution:

(a) 3.2

(b) If the distance increases by 3.2, then the force will decrease by a factor of 10 ; if the distance decreases by 3.2, then the force will increase by a factor of 10. Either way, the force changes by a factor of 10.

Exercise:

Problem:

Suppose you have a total charge q_{tot} that you can split in any manner. Once split, the separation distance is fixed. How do you split the charge to achieve the greatest force?

Exercise:

Problem:

(a) Common transparent tape becomes charged when pulled from a dispenser. If one piece is placed above another, the repulsive force can be great enough to support the top piece's weight. Assuming equal point charges (only an approximation), calculate the magnitude of the charge if electrostatic force is great enough to support the weight of a 10.0 mg piece of tape held 1.00 cm above another. (b) Discuss whether the magnitude of this charge is consistent with what is typical of static electricity.

Solution:

(a) $1.04 \times 10^{-9} \text{ C}$

(b) This charge is approximately 1 nC, which is consistent with the magnitude of charge typical for static electricity

Exercise:

Problem:

(a) Find the ratio of the electrostatic to gravitational force between two electrons. (b) What is this ratio for two protons? (c) Why is the ratio different for electrons and protons?

Exercise:

Problem:

At what distance is the electrostatic force between two protons equal to the weight of one proton?

Exercise:

Problem:

A certain five cent coin contains 5.00 g of nickel. What fraction of the nickel atoms' electrons, removed and placed 1.00 m above it, would support the weight of this coin? The atomic mass of nickel is 58.7, and each nickel atom contains 28 electrons and 28 protons.

Solution:

$$1.02 \times 10^{-11}$$

Exercise:**Problem:**

(a) Two point charges totaling $8.00 \mu\text{C}$ exert a repulsive force of 0.150 N on one another when separated by 0.500 m. What is the charge on each? (b) What is the charge on each if the force is attractive?

Exercise:**Problem:**

Point charges of $5.00 \mu\text{C}$ and $-3.00 \mu\text{C}$ are placed 0.250 m apart. (a) Where can a third charge be placed so that the net force on it is zero? (b) What if both charges are positive?

Solution:

- a. 0.859 m beyond negative charge on line connecting two charges
- b. 0.109 m from lesser charge on line connecting two charges

Exercise:**Problem:**

Two point charges q_1 and q_2 are 3.00 m apart, and their total charge is $20 \mu\text{C}$. (a) If the force of repulsion between them is 0.075N, what are magnitudes of the two charges? (b) If one charge attracts the other with a force of 0.525N, what are the magnitudes of the two charges? Note that you may need to solve a quadratic equation to reach your answer.

Glossary**Coulomb's law**

the mathematical equation calculating the electrostatic force vector between two charged particles

Coulomb force

another term for the electrostatic force

electrostatic force

the amount and direction of attraction or repulsion between two charged bodies

Electric Field: Concept of a Field Revisited

- Describe a force field and calculate the strength of an electric field due to a point charge.
- Calculate the force exerted on a test charge by an electric field.
- Explain the relationship between electrical force (F) on a test charge and electrical field strength (E).

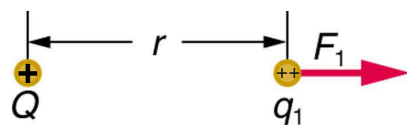
Contact forces, such as between a baseball and a bat, are explained on the small scale by the interaction of the charges in atoms and molecules in close proximity. They interact through forces that include the **Coulomb force**. Action at a distance is a force between objects that are not close enough for their atoms to “touch.” That is, they are separated by more than a few atomic diameters.

For example, a charged rubber comb attracts neutral bits of paper from a distance via the Coulomb force. It is very useful to think of an object being surrounded in space by a **force field**. The force field carries the force to another object (called a test object) some distance away.

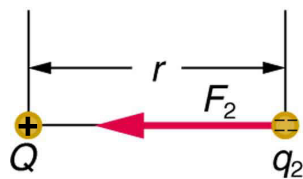
Concept of a Field

A field is a way of conceptualizing and mapping the force that surrounds any object and acts on another object at a distance without apparent physical connection. For example, the gravitational field surrounding the earth (and all other masses) represents the gravitational force that would be experienced if another mass were placed at a given point within the field.

In the same way, the Coulomb force field surrounding any charge extends throughout space. Using Coulomb’s law, $F = k|q_1q_2|/r^2$, its magnitude is given by the equation $F = k|qQ|/r^2$, for a **point charge** (a particle having a charge Q) acting on a **test charge** q at a distance r (see [\[link\]](#)). Both the magnitude and direction of the Coulomb force field depend on Q and the test charge q .



(a)



(b)

The Coulomb force field due to a positive charge Q is shown acting on two different charges. Both charges are the same distance from Q . (a) Since q_1 is positive, the force F_1 acting on it is repulsive. (b) The charge q_2 is negative and greater in magnitude than q_1 , and so the force F_2 acting on it is attractive and stronger than F_1 . The Coulomb force field is thus not unique at any point in space, because it depends on the test charges q_1 and q_2

as well as the
charge Q .

To simplify things, we would prefer to have a field that depends only on Q and not on the test charge q . The electric field is defined in such a manner that it represents only the charge creating it and is unique at every point in space. Specifically, the electric field E is defined to be the ratio of the Coulomb force to the test charge:

Equation:

$$\mathbf{E} = \frac{\mathbf{F}}{q},$$

where \mathbf{F} is the electrostatic force (or Coulomb force) exerted on a positive test charge q . It is understood that \mathbf{E} is in the same direction as \mathbf{F} . It is also assumed that q is so small that it does not alter the charge distribution creating the electric field. The units of electric field are newtons per coulomb (N/C). If the electric field is known, then the electrostatic force on any charge q is simply obtained by multiplying charge times electric field, or $\mathbf{F} = q\mathbf{E}$. Consider the electric field due to a point charge Q . According to Coulomb's law, the force it exerts on a test charge q is $F = k|qQ|/r^2$. Thus the magnitude of the electric field, E , for a point charge is

Equation:

$$E = \left| \frac{F}{q} \right| = k \left| \frac{qQ}{qr^2} \right| = k \frac{|Q|}{r^2}.$$

Since the test charge cancels, we see that

Equation:

$$E = k \frac{|Q|}{r^2}.$$

The electric field is thus seen to depend only on the charge Q and the distance r ; it is completely independent of the test charge q .

Example:

Calculating the Electric Field of a Point Charge

Calculate the strength and direction of the electric field E due to a point charge of 2.00 nC (nano-Coulombs) at a distance of 5.00 mm from the charge.

Strategy

We can find the electric field created by a point charge by using the equation $E = kQ/r^2$.

Solution

Here $Q = 2.00 \times 10^{-9} \text{ C}$ and $r = 5.00 \times 10^{-3} \text{ m}$. Entering those values into the above equation gives

Equation:

$$\begin{aligned} E &= k \frac{Q}{r^2} \\ &= (8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \times \frac{(2.00 \times 10^{-9} \text{ C})}{(5.00 \times 10^{-3} \text{ m})^2} \\ &= 7.19 \times 10^5 \text{ N/C}. \end{aligned}$$

Discussion

This **electric field strength** is the same at any point 5.00 mm away from the charge Q that creates the field. It is positive, meaning that it has a direction pointing away from the charge Q .

Example:

Calculating the Force Exerted on a Point Charge by an Electric Field

What force does the electric field found in the previous example exert on a point charge of $-0.250 \mu\text{C}$?

Strategy

Since we know the electric field strength and the charge in the field, the force on that charge can be calculated using the definition of electric field

$\mathbf{E} = \mathbf{F}/q$ rearranged to $\mathbf{F} = q\mathbf{E}$.

Solution

The magnitude of the force on a charge $q = -0.250 \mu\text{C}$ exerted by a field of strength $E = 7.20 \times 10^5 \text{ N/C}$ is thus,

Equation:

$$\begin{aligned} F &= -qE \\ &= (0.250 \times 10^{-6} \text{ C})(7.20 \times 10^5 \text{ N/C}) \\ &= 0.180 \text{ N.} \end{aligned}$$

Because q is negative, the force is directed opposite to the direction of the field.

Discussion

The force is attractive, as expected for unlike charges. (The field was created by a positive charge and here acts on a negative charge.) The charges in this example are typical of common static electricity, and the modest attractive force obtained is similar to forces experienced in static cling and similar situations.

Note:

PhET Explorations: Electric Field of Dreams

Play ball! Add charges to the Field of Dreams and see how they react to the electric field. Turn on a background electric field and adjust the direction and magnitude.

<https://archive.cnx.org/specials/ca9a78b4-06a7-11e6-b638-3bb71d1f0b42/electric-field-of-dreams/#sim-electric-field-of-dreams>

Section Summary

- The electrostatic force field surrounding a charged object extends out into space in all directions.
- The electrostatic force exerted by a point charge on a test charge at a distance r depends on the charge of both charges, as well as the

distance between the two.

- The electric field \mathbf{E} is defined to be
Equation:

$$\mathbf{E} = \frac{\mathbf{F}}{q},$$

where \mathbf{F} is the Coulomb or electrostatic force exerted on a small positive test charge q . \mathbf{E} has units of N/C.

- The magnitude of the electric field \mathbf{E} created by a point charge Q is
Equation:

$$\mathbf{E} = k \frac{|Q|}{r^2}.$$

where r is the distance from Q . The electric field \mathbf{E} is a vector and fields due to multiple charges add like vectors.

Conceptual Questions

Exercise:

Problem:

Why must the test charge q in the definition of the electric field be vanishingly small?

Exercise:

Problem:

Are the direction and magnitude of the Coulomb force unique at a given point in space? What about the electric field?

Problem Exercises

Exercise:

Problem:

What is the magnitude and direction of an electric field that exerts a 2.00×10^{-5} N upward force on a $-1.75 \mu\text{C}$ charge?

Exercise:**Problem:**

What is the magnitude and direction of the force exerted on a $3.50 \mu\text{C}$ charge by a 250 N/C electric field that points due east?

Solution:

$$8.75 \times 10^{-4} \text{ N}$$

Exercise:**Problem:**

Calculate the magnitude of the electric field 2.00 m from a point charge of 5.00 mC (such as found on the terminal of a Van de Graaff).

Exercise:**Problem:**

(a) What magnitude point charge creates a 10,000 N/C electric field at a distance of 0.250 m? (b) How large is the field at 10.0 m?

Solution:

(a) $6.94 \times 10^{-8} \text{ C}$

(b) 6.25 N/C

Exercise:

Problem:

Calculate the initial (from rest) acceleration of a proton in a $5.00 \times 10^6 \text{ N/C}$ electric field (such as created by a research Van de Graaff). Explicitly show how you follow the steps in the Problem-Solving Strategy for electrostatics.

Exercise:**Problem:**

(a) Find the magnitude and direction of an electric field that exerts a $4.80 \times 10^{-17} \text{ N}$ westward force on an electron. (b) What magnitude and direction force does this field exert on a proton?

Solution:

(a) 300 N/C (east)

(b) $4.80 \times 10^{-17} \text{ N}$ (east)

Glossary**field**

a map of the amount and direction of a force acting on other objects, extending out into space

point charge

A charged particle, designated Q , generating an electric field

test charge

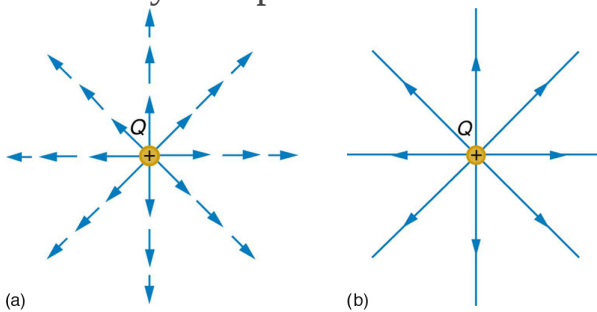
A particle (designated q) with either a positive or negative charge set down within an electric field generated by a point charge

Electric Field Lines: Multiple Charges

- Calculate the total force (magnitude and direction) exerted on a test charge from more than one charge
- Describe an electric field diagram of a positive point charge; of a negative point charge with twice the magnitude of positive charge
- Draw the electric field lines between two points of the same charge; between two points of opposite charge.

Drawings using lines to represent **electric fields** around charged objects are very useful in visualizing field strength and direction. Since the electric field has both magnitude and direction, it is a vector. Like all **vectors**, the electric field can be represented by an arrow that has length proportional to its magnitude and that points in the correct direction. (We have used arrows extensively to represent force vectors, for example.)

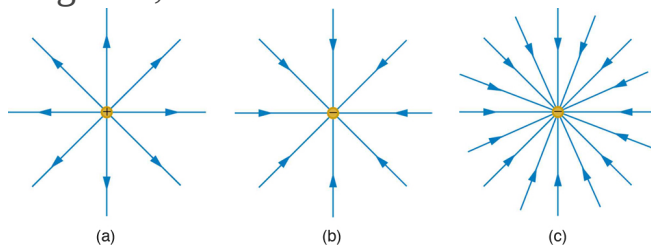
[\[link\]](#) shows two pictorial representations of the same electric field created by a positive point charge Q . [\[link\]](#) (b) shows the standard representation using continuous lines. [\[link\]](#) (a) shows numerous individual arrows with each arrow representing the force on a test charge q . Field lines are essentially a map of infinitesimal force vectors.



Two equivalent representations of the electric field due to a positive charge Q . (a) Arrows representing the electric field's magnitude and direction. (b) In the standard representation, the arrows are replaced by continuous field lines having the same direction at any point

as the electric field. The closeness of the lines is directly related to the strength of the electric field. A test charge placed anywhere will feel a force in the direction of the field line; this force will have a strength proportional to the density of the lines (being greater near the charge, for example).

Note that the electric field is defined for a positive test charge q , so that the field lines point away from a positive charge and toward a negative charge. (See [\[link\]](#).) The electric field strength is exactly proportional to the number of field lines per unit area, since the magnitude of the electric field for a point charge is $E = k|Q|/r^2$ and area is proportional to r^2 . This pictorial representation, in which field lines represent the direction and their closeness (that is, their areal density or the number of lines crossing a unit area) represents strength, is used for all fields: electrostatic, gravitational, magnetic, and others.



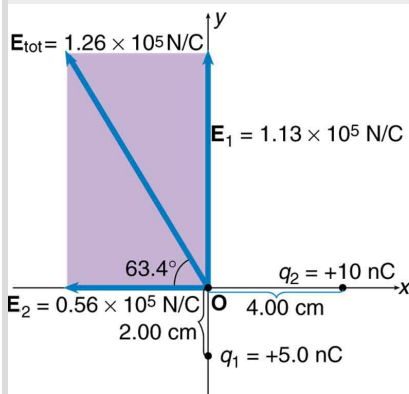
The electric field surrounding three different point charges. (a) A positive charge. (b) A negative charge of equal magnitude. (c) A larger negative charge.

In many situations, there are multiple charges. The total electric field created by multiple charges is the vector sum of the individual fields created by each charge. The following example shows how to add electric field vectors.

Example:

Adding Electric Fields

Find the magnitude and direction of the total electric field due to the two point charges, q_1 and q_2 , at the origin of the coordinate system as shown in [\[link\]](#).



The electric fields
 \mathbf{E}_1 and \mathbf{E}_2 at the
origin O add to
 \mathbf{E}_{tot} .

Strategy

Since the electric field is a vector (having magnitude and direction), we add electric fields with the same vector techniques used for other types of vectors. We first must find the electric field due to each charge at the point of interest, which is the origin of the coordinate system (O) in this instance. We pretend that there is a positive test charge, q , at point O , which allows us to determine the direction of the fields \mathbf{E}_1 and \mathbf{E}_2 . Once those fields are found, the total field can be determined using **vector addition**.

Solution

The electric field strength at the origin due to q_1 is labeled E_1 and is calculated:

Equation:

$$E_1 = k \frac{q_1}{r_1^2} = \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2 / \text{C}^2 \right) \frac{(5.00 \times 10^{-9} \text{ C})}{(2.00 \times 10^{-2} \text{ m})^2}$$
$$E_1 = 1.124 \times 10^5 \text{ N/C}.$$

Similarly, E_2 is

Equation:

$$E_2 = k \frac{q_2}{r_2^2} = \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2 / \text{C}^2 \right) \frac{(10.0 \times 10^{-9} \text{ C})}{(4.00 \times 10^{-2} \text{ m})^2}$$
$$E_2 = 0.5619 \times 10^5 \text{ N/C}.$$

Four digits have been retained in this solution to illustrate that E_1 is exactly twice the magnitude of E_2 . Now arrows are drawn to represent the magnitudes and directions of \mathbf{E}_1 and \mathbf{E}_2 . (See [\[link\]](#).) The direction of the electric field is that of the force on a positive charge so both arrows point directly away from the positive charges that create them. The arrow for \mathbf{E}_1 is exactly twice the length of that for \mathbf{E}_2 . The arrows form a right triangle in this case and can be added using the Pythagorean theorem. The magnitude of the total field E_{tot} is

Equation:

$$E_{\text{tot}} = (E_1^2 + E_2^2)^{1/2}$$
$$= \{(1.124 \times 10^5 \text{ N/C})^2 + (0.5619 \times 10^5 \text{ N/C})^2\}^{1/2}$$
$$= 1.26 \times 10^5 \text{ N/C}.$$

The direction is

Equation:

$$\begin{aligned}
 \theta &= \tan^{-1} \left(\frac{E_1}{E_2} \right) \\
 &= \tan^{-1} \left(\frac{1.124 \times 10^5 \text{ N/C}}{0.5619 \times 10^5 \text{ N/C}} \right) \\
 &= 63.4^\circ,
 \end{aligned}$$

or 63.4° above the x -axis.

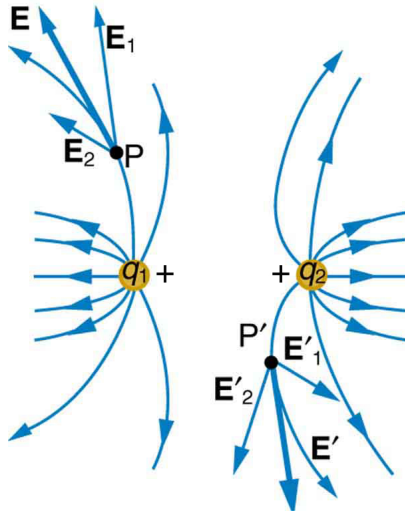
Discussion

In cases where the electric field vectors to be added are not perpendicular, vector components or graphical techniques can be used. The total electric field found in this example is the total electric field at only one point in space. To find the total electric field due to these two charges over an entire region, the same technique must be repeated for each point in the region. This impossibly lengthy task (there are an infinite number of points in space) can be avoided by calculating the total field at representative points and using some of the unifying features noted next.

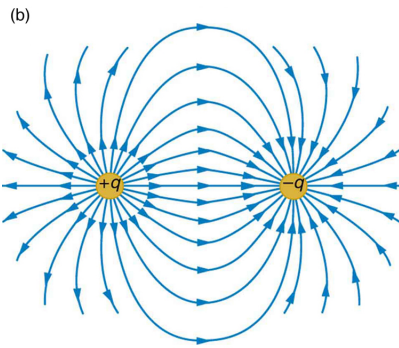
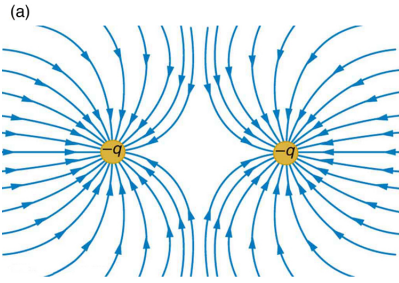
[\[link\]](#) shows how the electric field from two point charges can be drawn by finding the total field at representative points and drawing electric field lines consistent with those points. While the electric fields from multiple charges are more complex than those of single charges, some simple features are easily noticed.

For example, the field is weaker between like charges, as shown by the lines being farther apart in that region. (This is because the fields from each charge exert opposing forces on any charge placed between them.) (See [\[link\]](#) and [\[link\]](#)(a).) Furthermore, at a great distance from two like charges, the field becomes identical to the field from a single, larger charge.

[\[link\]](#)(b) shows the electric field of two unlike charges. The field is stronger between the charges. In that region, the fields from each charge are in the same direction, and so their strengths add. The field of two unlike charges is weak at large distances, because the fields of the individual charges are in opposite directions and so their strengths subtract. At very large distances, the field of two unlike charges looks like that of a smaller single charge.



Two positive point charges q_1 and q_2 produce the resultant electric field shown. The field is calculated at representative points and then smooth field lines drawn following the rules outlined in the text.



(a) Two negative charges produce the fields shown. It is very similar to the field produced by two positive charges, except that the directions are reversed. The field is clearly weaker between the charges. The individual forces on a test charge in that region are in opposite directions.

(b) Two opposite charges produce the field shown, which is stronger in the region between the charges.

We use electric field lines to visualize and analyze electric fields (the lines are a pictorial tool, not a physical entity in themselves). The properties of electric field lines for any charge distribution can be summarized as follows:

1. Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
2. The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
3. The strength of the field is proportional to the closeness of the field lines—more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
4. The direction of the electric field is tangent to the field line at any point in space.
5. Field lines can never cross.

The last property means that the field is unique at any point. The field line represents the direction of the field; so if they crossed, the field would have two directions at that location (an impossibility if the field is unique).

Note:

PhET Explorations: Charges and Fields

Move point charges around on the playing field and then view the electric field, voltages, equipotential lines, and more. It's colorful, it's dynamic, it's free.

[Click here
for the
simulation](#)

.

Section Summary

- Drawings of electric field lines are useful visual tools. The properties of electric field lines for any charge distribution are that:
- Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
- The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
- The strength of the field is proportional to the closeness of the field lines—more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
- The direction of the electric field is tangent to the field line at any point in space.
- Field lines can never cross.

Conceptual Questions

Exercise:

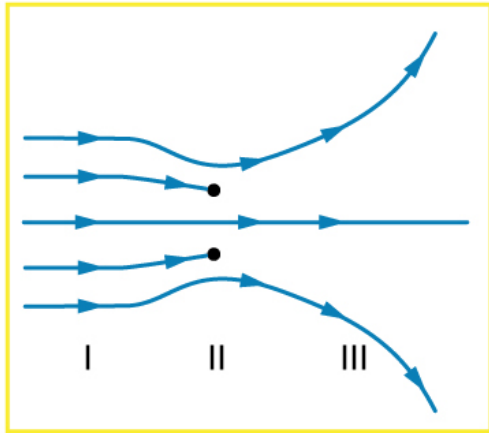
Problem:

Compare and contrast the Coulomb force field and the electric field. To do this, make a list of five properties for the Coulomb force field analogous to the five properties listed for electric field lines. Compare each item in your list of Coulomb force field properties with those of the electric field—are they the same or different? (For example, electric field lines cannot cross. Is the same true for Coulomb field lines?)

Exercise:

Problem:

[\[link\]](#) shows an electric field extending over three regions, labeled I, II, and III. Answer the following questions. (a) Are there any isolated charges? If so, in what region and what are their signs? (b) Where is the field strongest? (c) Where is it weakest? (d) Where is the field the most uniform?



Problem Exercises

Exercise:

Problem:

(a) Sketch the electric field lines near a point charge $+q$. (b) Do the same for a point charge $-3.00q$.

Exercise:

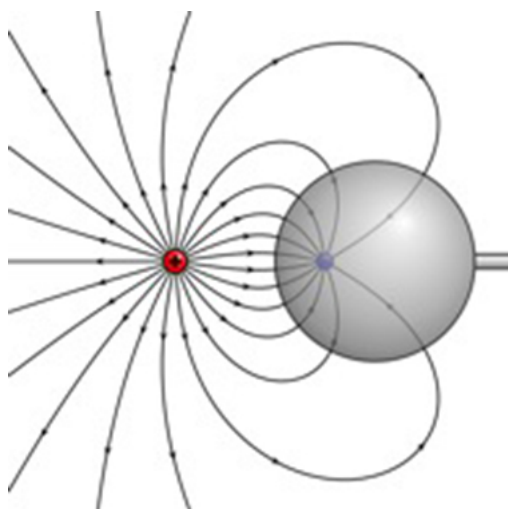
Problem:

Sketch the electric field lines a long distance from the charge distributions shown in [\[link\]](#) (a) and (b)

Exercise:

Problem:

[\[link\]](#) shows the electric field lines near two charges q_1 and q_2 . What is the ratio of their magnitudes? (b) Sketch the electric field lines a long distance from the charges shown in the figure.



The electric field near two charges.

Exercise:

Problem:

Sketch the electric field lines in the vicinity of two opposite charges, where the negative charge is three times greater in magnitude than the positive. (See [\[link\]](#) for a similar situation).

Glossary

electric field

a three-dimensional map of the electric force extended out into space from a point charge

electric field lines

a series of lines drawn from a point charge representing the magnitude and direction of force exerted by that charge

vector

a quantity with both magnitude and direction

vector addition

mathematical combination of two or more vectors, including their magnitudes, directions, and positions

Applications of Electrostatics

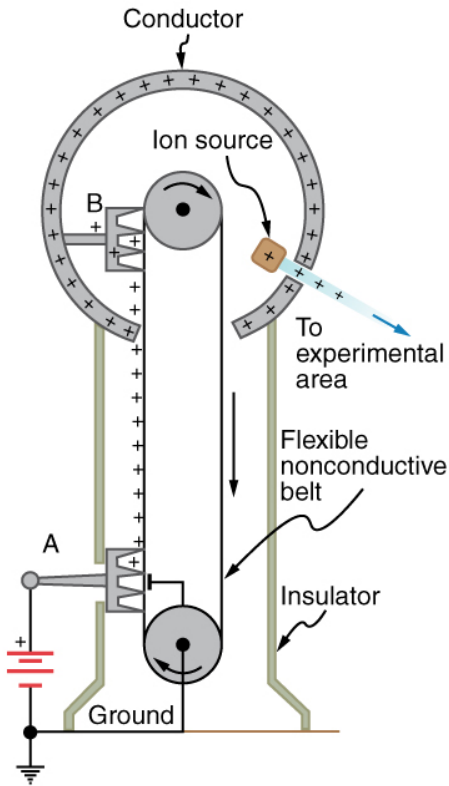
- Name several real-world applications of the study of electrostatics.

The study of **electrostatics** has proven useful in many areas. This module covers just a few of the many applications of electrostatics.

The Van de Graaff Generator

Van de Graaff generators (or Van de Graaffs) are not only spectacular devices used to demonstrate high voltage due to static electricity—they are also used for serious research. The first was built by Robert Van de Graaff in 1931 (based on original suggestions by Lord Kelvin) for use in nuclear physics research. [\[link\]](#) shows a schematic of a large research version. Van de Graaffs utilize both smooth and pointed surfaces, and conductors and insulators to generate large static charges and, hence, large voltages.

A very large excess charge can be deposited on the sphere, because it moves quickly to the outer surface. Practical limits arise because the large electric fields polarize and eventually ionize surrounding materials, creating free charges that neutralize excess charge or allow it to escape. Nevertheless, voltages of 15 million volts are well within practical limits.



Schematic of Van de Graaff generator. A battery (A) supplies excess positive charge to a pointed conductor, the points of which spray the charge onto a moving insulating belt near the bottom. The pointed conductor (B) on top in the large sphere picks up the charge. (The induced electric field at the points is so large that it removes the charge from the belt.) This can be done because the charge does not

remain inside the conducting sphere but moves to its outside surface. An ion source inside the sphere produces positive ions, which are accelerated away from the positive sphere to high velocities.

Note:

Take-Home Experiment: Electrostatics and Humidity

Rub a comb through your hair and use it to lift pieces of paper. It may help to tear the pieces of paper rather than cut them neatly. Repeat the exercise in your bathroom after you have had a long shower and the air in the bathroom is moist. Is it easier to get electrostatic effects in dry or moist air? Why would torn paper be more attractive to the comb than cut paper? Explain your observations.

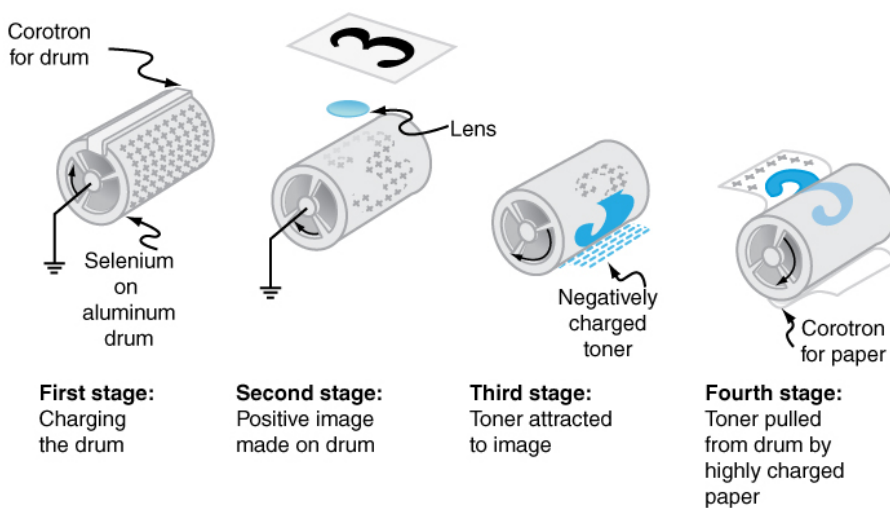
Xerography

Most copy machines use an electrostatic process called **xerography**—a word coined from the Greek words *xeros* for dry and *graphos* for writing. The heart of the process is shown in simplified form in [\[link\]](#).

A selenium-coated aluminum drum is sprayed with positive charge from points on a device called a corotron. Selenium is a substance with an interesting property—it is a **photoconductor**. That is, selenium is an insulator when in the dark and a conductor when exposed to light.

In the first stage of the xerography process, the conducting aluminum drum is **grounded** so that a negative charge is induced under the thin layer of uniformly positively charged selenium. In the second stage, the surface of the drum is exposed to the image of whatever is to be copied. Where the image is light, the selenium becomes conducting, and the positive charge is neutralized. In dark areas, the positive charge remains, and so the image has been transferred to the drum.

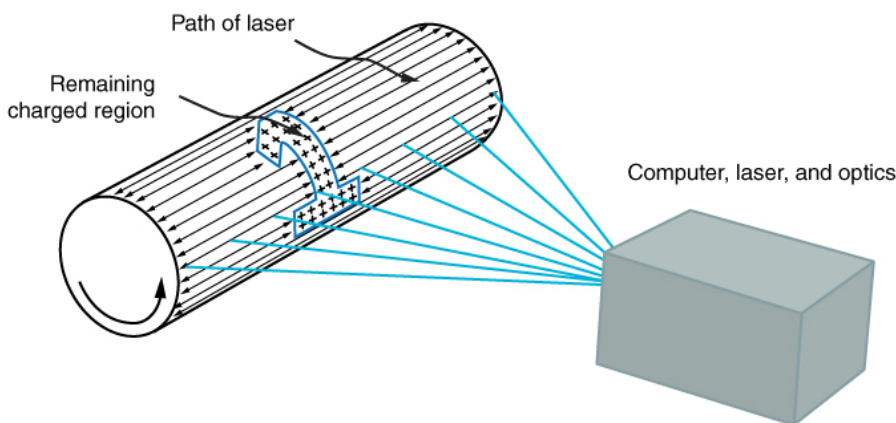
The third stage takes a dry black powder, called toner, and sprays it with a negative charge so that it will be attracted to the positive regions of the drum. Next, a blank piece of paper is given a greater positive charge than on the drum so that it will pull the toner from the drum. Finally, the paper and electrostatically held toner are passed through heated pressure rollers, which melt and permanently adhere the toner within the fibers of the paper.



Xerography is a dry copying process based on electrostatics. The major steps in the process are the charging of the photoconducting drum, transfer of an image creating a positive charge duplicate, attraction of toner to the charged parts of the drum, and transfer of toner to the paper. Not shown are heat treatment of the paper and cleansing of the drum for the next copy.

Laser Printers

Laser printers use the xerographic process to make high-quality images on paper, employing a laser to produce an image on the photoconducting drum as shown in [\[link\]](#). In its most common application, the laser printer receives output from a computer, and it can achieve high-quality output because of the precision with which laser light can be controlled. Many laser printers do significant information processing, such as making sophisticated letters or fonts, and may contain a computer more powerful than the one giving them the raw data to be printed.

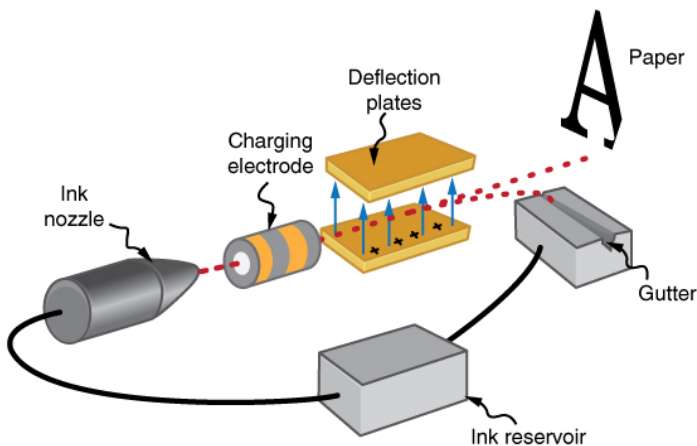


In a laser printer, a laser beam is scanned across a photoconducting drum, leaving a positive charge image. The other steps for charging the drum and transferring the image to paper are the same as in xerography. Laser light can be very precisely controlled, enabling laser printers to produce high-quality images.

Ink Jet Printers and Electrostatic Painting

The **ink jet printer**, commonly used to print computer-generated text and graphics, also employs electrostatics. A nozzle makes a fine spray of tiny ink droplets, which are then given an electrostatic charge. (See [\[link\]](#).)

Once charged, the droplets can be directed, using pairs of charged plates, with great precision to form letters and images on paper. Ink jet printers can produce color images by using a black jet and three other jets with primary colors, usually cyan, magenta, and yellow, much as a color television produces color. (This is more difficult with xerography, requiring multiple drums and toners.)



The nozzle of an ink-jet printer produces small ink droplets, which are sprayed with electrostatic charge. Various computer-driven devices are then used to direct the droplets to the correct positions on a page.

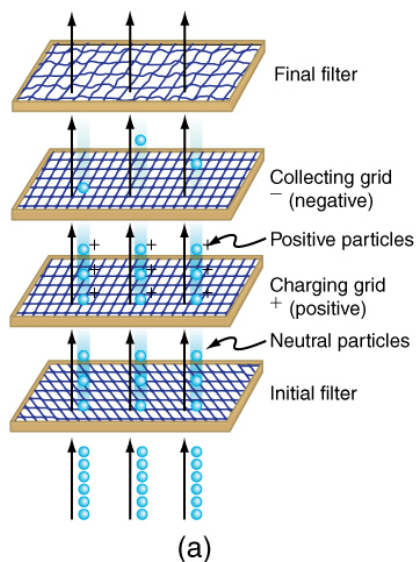
Electrostatic painting employs electrostatic charge to spray paint onto odd-shaped surfaces. Mutual repulsion of like charges causes the paint to fly away from its source. Surface tension forms drops, which are then attracted by unlike charges to the surface to be painted. Electrostatic painting can reach those hard-to-get at places, applying an even coat in a controlled

manner. If the object is a conductor, the electric field is perpendicular to the surface, tending to bring the drops in perpendicularly. Corners and points on conductors will receive extra paint. Felt can similarly be applied.

Smoke Precipitators and Electrostatic Air Cleaning

Another important application of electrostatics is found in air cleaners, both large and small. The electrostatic part of the process places excess (usually positive) charge on smoke, dust, pollen, and other particles in the air and then passes the air through an oppositely charged grid that attracts and retains the charged particles. (See [\[link\]](#).)

Large **electrostatic precipitators** are used industrially to remove over 99% of the particles from stack gas emissions associated with the burning of coal and oil. Home precipitators, often in conjunction with the home heating and air conditioning system, are very effective in removing polluting particles, irritants, and allergens.



(a) Schematic of an electrostatic precipitator. Air is passed through grids of opposite charge. The first grid charges airborne particles, while the second attracts and collects them. (b) The dramatic effect of

electrostatic precipitators is seen by the absence of smoke from this power plant. (credit: Cmdalgleish, Wikimedia Commons)

Note:

Problem-Solving Strategies for Electrostatics

1. Examine the situation to determine if static electricity is involved. This may concern separated stationary charges, the forces among them, and the electric fields they create.
2. Identify the system of interest. This includes noting the number, locations, and types of charges involved.
3. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful. Determine whether the Coulomb force is to be considered directly—if so, it may be useful to draw a free-body diagram, using electric field lines.
4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns). It is important to distinguish the Coulomb force F from the electric field E , for example.
5. Solve the appropriate equation for the quantity to be determined (the unknown) or draw the field lines as requested.
6. Examine the answer to see if it is reasonable: Does it make sense? Are units correct and the numbers involved reasonable?

Integrated Concepts

The Integrated Concepts exercises for this module involve concepts such as electric charges, electric fields, and several other topics. Physics is most interesting when applied to general situations involving more than a narrow set of physical principles. The electric field exerts force on charges, for example, and hence the relevance of [Dynamics: Force and Newton's Laws of Motion](#). The following topics are involved in some or all of the problems labeled “Integrated Concepts”:

- [Kinematics](#)
- [Two-Dimensional Kinematics](#)
- [Dynamics: Force and Newton's Laws of Motion](#)
- [Uniform Circular Motion and Gravitation](#)
- [Statics and Torque](#)
- [Fluid Statics](#)

The following worked example illustrates how this strategy is applied to an Integrated Concept problem:

Example:

Acceleration of a Charged Drop of Gasoline

If steps are not taken to ground a gasoline pump, static electricity can be placed on gasoline when filling your car's tank. Suppose a tiny drop of gasoline has a mass of 4.00×10^{-15} kg and is given a positive charge of 3.20×10^{-19} C. (a) Find the weight of the drop. (b) Calculate the electric force on the drop if there is an upward electric field of strength 3.00×10^5 N/C due to other static electricity in the vicinity. (c) Calculate the drop's acceleration.

Strategy

To solve an integrated concept problem, we must first identify the physical principles involved and identify the chapters in which they are found. Part (a) of this example asks for weight. This is a topic of dynamics and is defined in [Dynamics: Force and Newton's Laws of Motion](#). Part (b) deals with electric force on a charge, a topic of [Electric Charge and Electric Field](#). Part (c) asks for acceleration, knowing forces and mass. These are part of Newton's laws, also found in [Dynamics: Force and Newton's Laws of Motion](#).

The following solutions to each part of the example illustrate how the specific problem-solving strategies are applied. These involve identifying knowns and unknowns, checking to see if the answer is reasonable, and so on.

Solution for (a)

Weight is mass times the acceleration due to gravity, as first expressed in

Equation:

$$w = mg.$$

Entering the given mass and the average acceleration due to gravity yields

Equation:

$$w = (4.00 \times 10^{-15} \text{ kg})(9.80 \text{ m/s}^2) = 3.92 \times 10^{-14} \text{ N}.$$

Discussion for (a)

This is a small weight, consistent with the small mass of the drop.

Solution for (b)

The force an electric field exerts on a charge is given by rearranging the following equation:

Equation:

$$F = qE.$$

Here we are given the charge ($3.20 \times 10^{-19} \text{ C}$ is twice the fundamental unit of charge) and the electric field strength, and so the electric force is found to be

Equation:

$$F = (3.20 \times 10^{-19} \text{ C})(3.00 \times 10^5 \text{ N/C}) = 9.60 \times 10^{-14} \text{ N}.$$

Discussion for (b)

While this is a small force, it is greater than the weight of the drop.

Solution for (c)

The acceleration can be found using Newton's second law, provided we can identify all of the external forces acting on the drop. We assume only the drop's weight and the electric force are significant. Since the drop has a positive charge and the electric field is given to be upward, the electric force is upward. We thus have a one-dimensional (vertical direction) problem, and we can state Newton's second law as

Equation:

$$a = \frac{F_{\text{net}}}{m}.$$

where $F_{\text{net}} = F - w$. Entering this and the known values into the expression for Newton's second law yields

Equation:

$$\begin{aligned} a &= \frac{F-w}{m} \\ &= \frac{9.60 \times 10^{-14} \text{ N} - 3.92 \times 10^{-14} \text{ N}}{4.00 \times 10^{-15} \text{ kg}} \\ &= 14.2 \text{ m/s}^2. \end{aligned}$$

Discussion for (c)

This is an upward acceleration great enough to carry the drop to places where you might not wish to have gasoline.

This worked example illustrates how to apply problem-solving strategies to situations that include topics in different chapters. The first step is to identify the physical principles involved in the problem. The second step is to solve for the unknown using familiar problem-solving strategies. These are found throughout the text, and many worked examples show how to use them for single topics. In this integrated concepts example, you can see how to apply them across several topics. You will find these techniques useful in applications of physics outside a physics course, such as in your profession, in other science disciplines, and in everyday life. The following problems will build your skills in the broad application of physical principles.

Note:

Unreasonable Results

The Unreasonable Results exercises for this module have results that are unreasonable because some premise is unreasonable or because certain of the premises are inconsistent with one another. Physical principles applied correctly then produce unreasonable results. The purpose of these problems is to give practice in assessing whether nature is being accurately described, and if it is not to trace the source of difficulty.

Note:**Problem-Solving Strategy**

To determine if an answer is reasonable, and to determine the cause if it is not, do the following.

1. Solve the problem using strategies as outlined above. Use the format followed in the worked examples in the text to solve the problem as usual.
2. Check to see if the answer is reasonable. Is it too large or too small, or does it have the wrong sign, improper units, and so on?
3. If the answer is unreasonable, look for what specifically could cause the identified difficulty. Usually, the manner in which the answer is unreasonable is an indication of the difficulty. For example, an extremely large Coulomb force could be due to the assumption of an excessively large separated charge.

Section Summary

- Electrostatics is the study of electric fields in static equilibrium.
- In addition to research using equipment such as a Van de Graaff generator, many practical applications of electrostatics exist, including photocopiers, laser printers, ink-jet printers and electrostatic air filters.

Problems & Exercises**Exercise:**

Problem:

(a) What is the electric field 5.00 m from the center of the terminal of a Van de Graaff with a 3.00 mC charge, noting that the field is equivalent to that of a point charge at the center of the terminal? (b) At this distance, what force does the field exert on a $2.00\ \mu\text{C}$ charge on the Van de Graaff's belt?

Exercise:**Problem:**

(a) What is the direction and magnitude of an electric field that supports the weight of a free electron near the surface of Earth? (b) Discuss what the small value for this field implies regarding the relative strength of the gravitational and electrostatic forces.

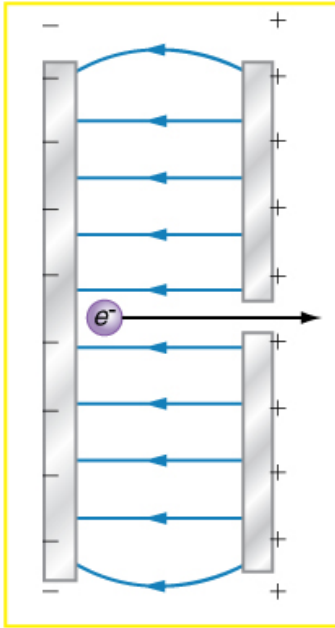
Solution:

(a) $5.58 \times 10^{-11}\ \text{N/C}$

(b) the coulomb force is extraordinarily stronger than gravity

Exercise:**Problem:**

A simple and common technique for accelerating electrons is shown in [\[link\]](#), where there is a uniform electric field between two plates. Electrons are released, usually from a hot filament, near the negative plate, and there is a small hole in the positive plate that allows the electrons to continue moving. (a) Calculate the acceleration of the electron if the field strength is $2.50 \times 10^4\ \text{N/C}$. (b) Explain why the electron will not be pulled back to the positive plate once it moves through the hole.



Parallel
conducting
plates with
opposite charges
on them create a
relatively
uniform electric
field used to
accelerate
electrons to the
right. Those that
go through the
hole can be used
to make a TV or
computer screen
glow or to
produce X-rays.

Exercise:

Problem:

Earth has a net charge that produces an electric field of approximately 150 N/C downward at its surface. (a) What is the magnitude and sign of the excess charge, noting the electric field of a conducting sphere is equivalent to a point charge at its center? (b) What acceleration will the field produce on a free electron near Earth's surface? (c) What mass object with a single extra electron will have its weight supported by this field?

Solution:

(a) $-6.76 \times 10^5 \text{ C}$

(b) $2.63 \times 10^{13} \text{ m/s}^2$ (upward)

(c) $2.45 \times 10^{-18} \text{ kg}$

Exercise:**Problem:**

Point charges of $25.0 \mu\text{C}$ and $45.0 \mu\text{C}$ are placed 0.500 m apart. (a) At what point along the line between them is the electric field zero? (b) What is the electric field halfway between them?

Exercise:**Problem:**

What can you say about two charges q_1 and q_2 , if the electric field one-fourth of the way from q_1 to q_2 is zero?

Solution:

The charge q_2 is 9 times greater than q_1 .

Exercise:**Problem: Integrated Concepts**

Calculate the angular velocity ω of an electron orbiting a proton in the hydrogen atom, given the radius of the orbit is 0.530×10^{-10} m. You may assume that the proton is stationary and the centripetal force is supplied by Coulomb attraction.

Exercise:

Problem: Integrated Concepts

An electron has an initial velocity of 5.00×10^6 m/s in a uniform 2.00×10^5 N/C strength electric field. The field accelerates the electron in the direction opposite to its initial velocity. (a) What is the direction of the electric field? (b) How far does the electron travel before coming to rest? (c) How long does it take the electron to come to rest? (d) What is the electron's velocity when it returns to its starting point?

Exercise:

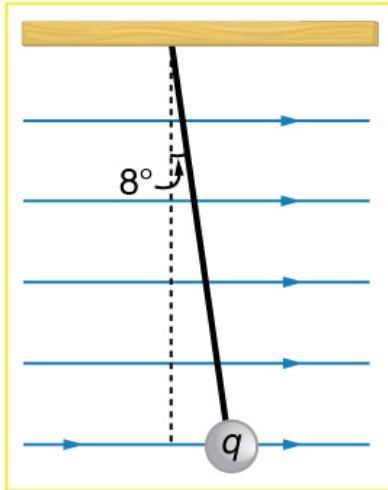
Problem: Integrated Concepts

The practical limit to an electric field in air is about 3.00×10^6 N/C. Above this strength, sparking takes place because air begins to ionize and charges flow, reducing the field. (a) Calculate the distance a free proton must travel in this field to reach 3.00% of the speed of light, starting from rest. (b) Is this practical in air, or must it occur in a vacuum?

Exercise:

Problem: Integrated Concepts

A 5.00 g charged insulating ball hangs on a 30.0 cm long string in a uniform horizontal electric field as shown in [\[link\]](#). Given the charge on the ball is $1.00 \mu\text{C}$, find the strength of the field.

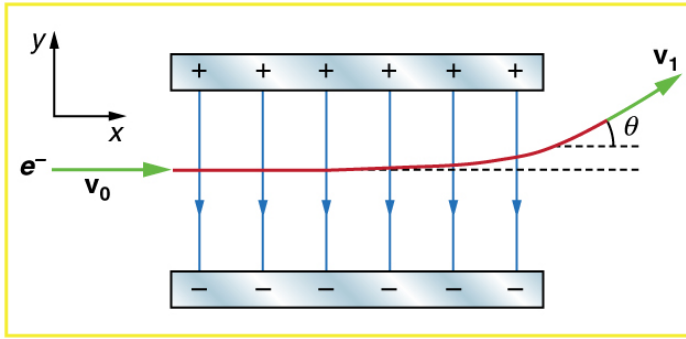


A horizontal electric field causes the charged ball to hang at an angle of 8.00° .

Exercise:

Problem: Integrated Concepts

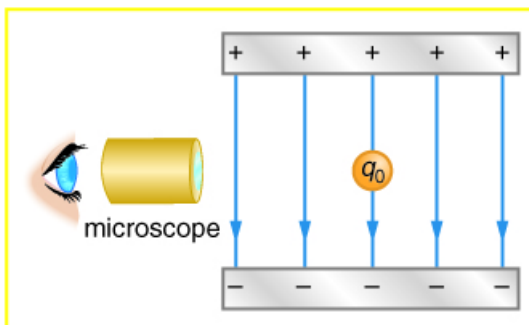
[\[link\]](#) shows an electron passing between two charged metal plates that create an 100 N/C vertical electric field perpendicular to the electron's original horizontal velocity. (These can be used to change the electron's direction, such as in an oscilloscope.) The initial speed of the electron is $3.00 \times 10^6 \text{ m/s}$, and the horizontal distance it travels in the uniform field is 4.00 cm . (a) What is its vertical deflection? (b) What is the vertical component of its final velocity? (c) At what angle does it exit? Neglect any edge effects.



Exercise:

Problem: Integrated Concepts

The classic Millikan oil drop experiment was the first to obtain an accurate measurement of the charge on an electron. In it, oil drops were suspended against the gravitational force by a vertical electric field. (See [\[link\]](#).) Given the oil drop to be $1.00 \mu\text{m}$ in radius and have a density of 920 kg/m^3 : (a) Find the weight of the drop. (b) If the drop has a single excess electron, find the electric field strength needed to balance its weight.



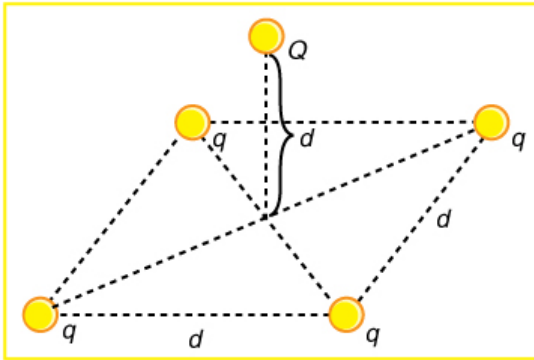
In the Millikan oil drop experiment, small drops can be suspended in an electric field by the force exerted on a single excess electron. Classically, this experiment was used to determine the electron charge q_e by

measuring the electric field
and mass of the drop.

Exercise:

Problem: Integrated Concepts

(a) In [\[link\]](#), four equal charges q lie on the corners of a square. A fifth charge Q is on a mass m directly above the center of the square, at a height equal to the length d of one side of the square. Determine the magnitude of q in terms of Q , m , and d , if the Coulomb force is to equal the weight of m . (b) Is this equilibrium stable or unstable? Discuss.



Four equal charges on the
corners of a horizontal
square support the weight of
a fifth charge located
directly above the center of
the square.

Exercise:

Problem: Unreasonable Results

(a) Calculate the electric field strength near a 10.0 cm diameter
conducting sphere that has 1.00 C of excess charge on it. (b) What is

unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Unreasonable Results

(a) Two 0.500 g raindrops in a thunderhead are 1.00 cm apart when they each acquire 1.00 mC charges. Find their acceleration. (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Exercise:

Problem: Unreasonable Results

A wrecking yard inventor wants to pick up cars by charging a 0.400 m diameter ball and inducing an equal and opposite charge on the car. If a car has a 1000 kg mass and the ball is to be able to lift it from a distance of 1.00 m: (a) What minimum charge must be used? (b) What is the electric field near the surface of the ball? (c) Why are these results unreasonable? (d) Which premise or assumption is responsible?

Exercise:

Problem: Construct Your Own Problem

Consider two insulating balls with evenly distributed equal and opposite charges on their surfaces, held with a certain distance between the centers of the balls. Construct a problem in which you calculate the electric field (magnitude and direction) due to the balls at various points along a line running through the centers of the balls and extending to infinity on either side. Choose interesting points and comment on the meaning of the field at those points. For example, at what points might the field be just that due to one ball and where does the field become negligibly small? Among the things to be considered are the magnitudes of the charges and the distance between the centers of the balls. Your instructor may wish for you to consider the electric

field off axis or for a more complex array of charges, such as those in a water molecule.

Exercise:

Problem: Construct Your Own Problem

Consider identical spherical conducting space ships in deep space where gravitational fields from other bodies are negligible compared to the gravitational attraction between the ships. Construct a problem in which you place identical excess charges on the space ships to exactly counter their gravitational attraction. Calculate the amount of excess charge needed. Examine whether that charge depends on the distance between the centers of the ships, the masses of the ships, or any other factors. Discuss whether this would be an easy, difficult, or even impossible thing to do in practice.

Glossary

Van de Graaff generator

a machine that produces a large amount of excess charge, used for experiments with high voltage

electrostatics

the study of electric forces that are static or slow-moving

photoconductor

a substance that is an insulator until it is exposed to light, when it becomes a conductor

xerography

a dry copying process based on electrostatics

grounded

connected to the ground with a conductor, so that charge flows freely to and from the Earth to the grounded object

laser printer

uses a laser to create a photoconductive image on a drum, which attracts dry ink particles that are then rolled onto a sheet of paper to print a high-quality copy of the image

ink-jet printer

small ink droplets sprayed with an electric charge are controlled by electrostatic plates to create images on paper

electrostatic precipitators

filters that apply charges to particles in the air, then attract those charges to a filter, removing them from the airstream

Introduction to Electric Potential and Electric Energy

class="introduction"

Automated
external
defibrillato
r unit
(AED)
(credit:
U.S.
Defense
Department
photo/Tech.
Sgt.
Suzanne
M. Day)



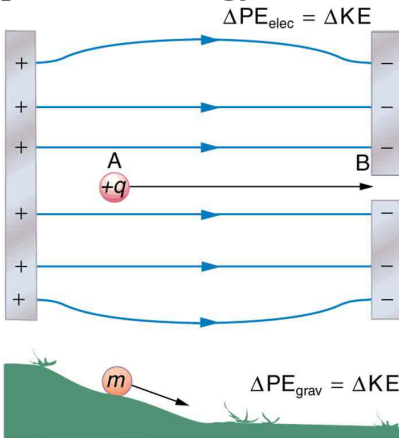
In [Electric Charge and Electric Field](#), we just scratched the surface (or at least rubbed it) of electrical phenomena. Two of the most familiar aspects of

electricity are its energy and *voltage*. We know, for example, that great amounts of electrical energy can be stored in batteries, are transmitted cross-country through power lines, and may jump from clouds to explode the sap of trees. In a similar manner, at molecular levels, *ions* cross cell membranes and transfer information. We also know about voltages associated with electricity. Batteries are typically a few volts, the outlets in your home produce 120 volts, and power lines can be as high as hundreds of thousands of volts. But energy and voltage are not the same thing. A motorcycle battery, for example, is small and would not be very successful in replacing the much larger car battery, yet each has the same voltage. In this chapter, we shall examine the relationship between voltage and electrical energy and begin to explore some of the many applications of electricity.

Electric Potential Energy: Potential Difference

- Define electric potential and electric potential energy.
- Describe the relationship between potential difference and electrical potential energy.
- Explain electron volt and its usage in submicroscopic process.
- Determine electric potential energy given potential difference and amount of charge.

When a free positive charge q is accelerated by an electric field, such as shown in [\[link\]](#), it is given kinetic energy. The process is analogous to an object being accelerated by a gravitational field. It is as if the charge is going down an electrical hill where its electric potential energy is converted to kinetic energy. Let us explore the work done on a charge q by the electric field in this process, so that we may develop a definition of electric potential energy.



A charge accelerated by an electric field is analogous to a mass going down a hill. In both cases potential energy is converted to another form. Work is done by a force, but since this force

is conservative, we
can write
 $W = -\Delta PE.$

The electrostatic or Coulomb force is conservative, which means that the work done on q is independent of the path taken. This is exactly analogous to the gravitational force in the absence of dissipative forces such as friction. When a force is conservative, it is possible to define a potential energy associated with the force, and it is usually easier to deal with the potential energy (because it depends only on position) than to calculate the work directly.

We use the letters PE to denote electric potential energy, which has units of joules (J). The change in potential energy, ΔPE , is crucial, since the work done by a conservative force is the negative of the change in potential energy; that is, $W = -\Delta PE$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE . There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Note:

Potential Energy

$W = -\Delta PE$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE . There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Gravitational potential energy and electric potential energy are quite analogous. Potential energy accounts for work done by a conservative force and gives added insight regarding energy and energy transformation

without the necessity of dealing with the force directly. It is much more common, for example, to use the concept of voltage (related to electric potential energy) than to deal with the Coulomb force directly.

Calculating the work directly is generally difficult, since $W = Fd \cos \theta$ and the direction and magnitude of F can be complex for multiple charges, for odd-shaped objects, and along arbitrary paths. But we do know that, since $F = qE$, the work, and hence ΔPE , is proportional to the test charge q . To have a physical quantity that is independent of test charge, we define **electric potential** V (or simply potential, since electric is understood) to be the potential energy per unit charge:

Equation:

$$V = \frac{PE}{q}.$$

Note:

Electric Potential

This is the electric potential energy per unit charge.

Equation:

$$V = \frac{PE}{q}$$

Since PE is proportional to q , the dependence on q cancels. Thus V does not depend on q . The change in potential energy ΔPE is crucial, and so we are concerned with the difference in potential or potential difference ΔV between two points, where

Equation:

$$\Delta V = V_B - V_A = \frac{\Delta PE}{q}.$$

The **potential difference** between points A and B, $V_B - V_A$, is thus defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

Equation:

$$1 \text{ V} = 1 \frac{\text{J}}{\text{C}}$$

Note:

Potential Difference

The potential difference between points A and B, $V_B - V_A$, is defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

Equation:

$$1 \text{ V} = 1 \frac{\text{J}}{\text{C}}$$

The familiar term **voltage** is the common name for potential difference. Keep in mind that whenever a voltage is quoted, it is understood to be the potential difference between two points. For example, every battery has two terminals, and its voltage is the potential difference between them. More fundamentally, the point you choose to be zero volts is arbitrary. This is analogous to the fact that gravitational potential energy has an arbitrary zero, such as sea level or perhaps a lecture hall floor.

In summary, the relationship between potential difference (or voltage) and electrical potential energy is given by

Equation:

$$\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V.$$

Note:**Potential Difference and Electrical Potential Energy**

The relationship between potential difference (or voltage) and electrical potential energy is given by

Equation:

$$\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V.$$

The second equation is equivalent to the first.

Voltage is not the same as energy. Voltage is the energy per unit charge. Thus a motorcycle battery and a car battery can both have the same voltage (more precisely, the same potential difference between battery terminals), yet one stores much more energy than the other since $\Delta \text{PE} = q\Delta V$. The car battery can move more charge than the motorcycle battery, although both are 12 V batteries.

Example:**Calculating Energy**

Suppose you have a 12.0 V motorcycle battery that can move 5000 C of charge, and a 12.0 V car battery that can move 60,000 C of charge. How much energy does each deliver? (Assume that the numerical value of each charge is accurate to three significant figures.)

Strategy

To say we have a 12.0 V battery means that its terminals have a 12.0 V potential difference. When such a battery moves charge, it puts the charge

through a potential difference of 12.0 V, and the charge is given a change in potential energy equal to $\Delta PE = q\Delta V$.

So to find the energy output, we multiply the charge moved by the potential difference.

Solution

For the motorcycle battery, $q = 5000 \text{ C}$ and $\Delta V = 12.0 \text{ V}$. The total energy delivered by the motorcycle battery is

Equation:

$$\begin{aligned}\Delta PE_{\text{cycle}} &= (5000 \text{ C})(12.0 \text{ V}) \\ &= (5000 \text{ C})(12.0 \text{ J/C}) \\ &= 6.00 \times 10^4 \text{ J}.\end{aligned}$$

Similarly, for the car battery, $q = 60,000 \text{ C}$ and

Equation:

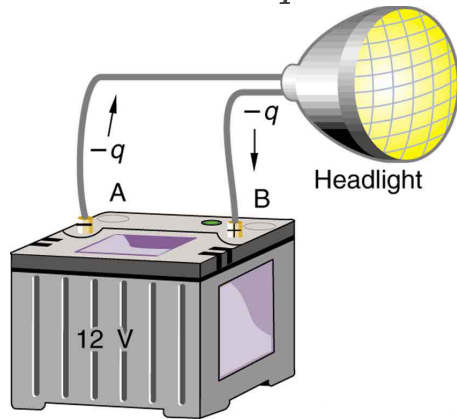
$$\begin{aligned}\Delta PE_{\text{car}} &= (60,000 \text{ C})(12.0 \text{ V}) \\ &= 7.20 \times 10^5 \text{ J}.\end{aligned}$$

Discussion

While voltage and energy are related, they are not the same thing. The voltages of the batteries are identical, but the energy supplied by each is quite different. Note also that as a battery is discharged, some of its energy is used internally and its terminal voltage drops, such as when headlights dim because of a low car battery. The energy supplied by the battery is still calculated as in this example, but not all of the energy is available for external use.

Note that the energies calculated in the previous example are absolute values. The change in potential energy for the battery is negative, since it loses energy. These batteries, like many electrical systems, actually move negative charge—electrons in particular. The batteries repel electrons from their negative terminals (A) through whatever circuitry is involved and attract them to their positive terminals (B) as shown in [\[link\]](#). The change in potential is $\Delta V = V_B - V_A = +12 \text{ V}$ and the charge q is negative, so that

$\Delta PE = q\Delta V$ is negative, meaning the potential energy of the battery has decreased when q has moved from A to B.



A battery moves negative charge from its negative terminal through a headlight to its positive terminal.

Appropriate combinations of chemicals in the battery separate charges so that the negative terminal has an excess of negative charge, which is repelled by it and attracted to the excess positive charge on the other terminal. In terms of potential, the positive terminal is at a higher voltage than the negative. Inside the battery, both positive and negative charges move.

Example:**How Many Electrons Move through a Headlight Each Second?**

When a 12.0 V car battery runs a single 30.0 W headlight, how many electrons pass through it each second?

Strategy

To find the number of electrons, we must first find the charge that moved in 1.00 s. The charge moved is related to voltage and energy through the equation $\Delta PE = q\Delta V$. A 30.0 W lamp uses 30.0 joules per second. Since the battery loses energy, we have $\Delta PE = -30.0$ J and, since the electrons are going from the negative terminal to the positive, we see that $\Delta V = +12.0$ V.

Solution

To find the charge q moved, we solve the equation $\Delta PE = q\Delta V$:

Equation:

$$q = \frac{\Delta PE}{\Delta V}.$$

Entering the values for ΔPE and ΔV , we get

Equation:

$$q = \frac{-30.0 \text{ J}}{+12.0 \text{ V}} = \frac{-30.0 \text{ J}}{+12.0 \text{ J/C}} = -2.50 \text{ C}.$$

The number of electrons n_e is the total charge divided by the charge per electron. That is,

Equation:

$$n_e = \frac{-2.50 \text{ C}}{-1.60 \times 10^{-19} \text{ C/e}^-} = 1.56 \times 10^{19} \text{ electrons}.$$

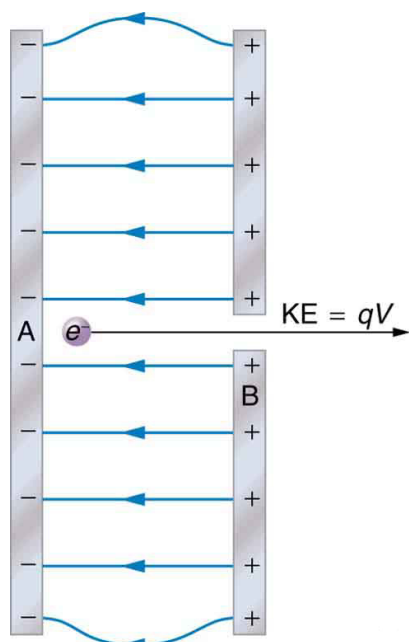
Discussion

This is a very large number. It is no wonder that we do not ordinarily observe individual electrons with so many being present in ordinary

systems. In fact, electricity had been in use for many decades before it was determined that the moving charges in many circumstances were negative. Positive charge moving in the opposite direction of negative charge often produces identical effects; this makes it difficult to determine which is moving or whether both are moving.

The Electron Volt

The energy per electron is very small in macroscopic situations like that in the previous example—a tiny fraction of a joule. But on a submicroscopic scale, such energy per particle (electron, proton, or ion) can be of great importance. For example, even a tiny fraction of a joule can be great enough for these particles to destroy organic molecules and harm living tissue. The particle may do its damage by direct collision, or it may create harmful x rays, which can also inflict damage. It is useful to have an energy unit related to submicroscopic effects. [\[link\]](#) shows a situation related to the definition of such an energy unit. An electron is accelerated between two charged metal plates as it might be in an old-model television tube or oscilloscope. The electron is given kinetic energy that is later converted to another form—light in the television tube, for example. (Note that downhill for the electron is uphill for a positive charge.) Since energy is related to voltage by $\Delta PE = q\Delta V$, we can think of the joule as a coulomb-volt.



A typical electron gun accelerates electrons using a potential difference between two metal plates. The energy of the electron in electron volts is numerically the same as the voltage between the plates.

For example, a 5000 V potential difference produces 5000 eV electrons.

On the submicroscopic scale, it is more convenient to define an energy unit called the **electron volt** (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form, **Equation:**

$$\begin{aligned}
 1 \text{ eV} &= (1.60 \times 10^{-19} \text{ C})(1 \text{ V}) = (1.60 \times 10^{-19} \text{ C})(1 \text{ J/C}) \\
 &= 1.60 \times 10^{-19} \text{ J}.
 \end{aligned}$$

Note:

Electron Volt

On the submicroscopic scale, it is more convenient to define an energy unit called the electron volt (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

Equation:

$$\begin{aligned}
 1 \text{ eV} &= (1.60 \times 10^{-19} \text{ C})(1 \text{ V}) = (1.60 \times 10^{-19} \text{ C})(1 \text{ J/C}) \\
 &= 1.60 \times 10^{-19} \text{ J}.
 \end{aligned}$$

An electron accelerated through a potential difference of 1 V is given an energy of 1 eV. It follows that an electron accelerated through 50 V is given 50 eV. A potential difference of 100,000 V (100 kV) will give an electron an energy of 100,000 eV (100 keV), and so on. Similarly, an ion with a double positive charge accelerated through 100 V will be given 200 eV of energy. These simple relationships between accelerating voltage and particle charges make the electron volt a simple and convenient energy unit in such circumstances.

Note:

Connections: Energy Units

The electron volt (eV) is the most common energy unit for submicroscopic processes. This will be particularly noticeable in the chapters on modern physics. Energy is so important to so many subjects that there is a tendency to define a special energy unit for each major topic. There are, for example,

calories for food energy, kilowatt-hours for electrical energy, and therms for natural gas energy.

The electron volt is commonly employed in submicroscopic processes—chemical valence energies and molecular and nuclear binding energies are among the quantities often expressed in electron volts. For example, about 5 eV of energy is required to break up certain organic molecules. If a proton is accelerated from rest through a potential difference of 30 kV, it is given an energy of 30 keV (30,000 eV) and it can break up as many as 6000 of these molecules ($30,000 \text{ eV} \div 5 \text{ eV per molecule} = 6000 \text{ molecules}$). Nuclear decay energies are on the order of 1 MeV (1,000,000 eV) per event and can, thus, produce significant biological damage.

Conservation of Energy

The total energy of a system is conserved if there is no net addition (or subtraction) of work or heat transfer. For conservative forces, such as the electrostatic force, conservation of energy states that mechanical energy is a constant.

Mechanical energy is the sum of the kinetic energy and potential energy of a system; that is, $KE + PE = \text{constant}$. A loss of PE of a charged particle becomes an increase in its KE. Here PE is the electric potential energy.

Conservation of energy is stated in equation form as

Equation:

$$KE + PE = \text{constant}$$

or

Equation:

$$KE_i + PE_i = KE_f + PE_f,$$

where i and f stand for initial and final conditions. As we have found many times before, considering energy can give us insights and facilitate problem solving.

Example:

Electrical Potential Energy Converted to Kinetic Energy

Calculate the final speed of a free electron accelerated from rest through a potential difference of 100 V. (Assume that this numerical value is accurate to three significant figures.)

Strategy

We have a system with only conservative forces. Assuming the electron is accelerated in a vacuum, and neglecting the gravitational force (we will check on this assumption later), all of the electrical potential energy is converted into kinetic energy. We can identify the initial and final forms of energy to be $KE_i = 0$, $KE_f = \frac{1}{2}mv^2$, $PE_i = qV$, and $PE_f = 0$.

Solution

Conservation of energy states that

Equation:

$$KE_i + PE_i = KE_f + PE_f.$$

Entering the forms identified above, we obtain

Equation:

$$qV = \frac{mv^2}{2}.$$

We solve this for v :

Equation:

$$v = \sqrt{\frac{2qV}{m}}.$$

Entering values for q , V , and m gives

Equation:

$$\begin{aligned}
 v &= \sqrt{\frac{2(-1.60 \times 10^{-19} \text{ C})(-100 \text{ J/C})}{9.11 \times 10^{-31} \text{ kg}}} \\
 &= 5.93 \times 10^6 \text{ m/s.}
 \end{aligned}$$

Discussion

Note that both the charge and the initial voltage are negative, as in [\[link\]](#). From the discussions in [Electric Charge and Electric Field](#), we know that electrostatic forces on small particles are generally very large compared with the gravitational force. The large final speed confirms that the gravitational force is indeed negligible here. The large speed also indicates how easy it is to accelerate electrons with small voltages because of their very small mass. Voltages much higher than the 100 V in this problem are typically used in electron guns. Those higher voltages produce electron speeds so great that relativistic effects must be taken into account. That is why a low voltage is considered (accurately) in this example.

Section Summary

- Electric potential is potential energy per unit charge.
- The potential difference between points A and B, $V_B - V_A$, defined to be the change in potential energy of a charge q moved from A to B, is equal to the change in potential energy divided by the charge. Potential difference is commonly called voltage, represented by the symbol ΔV .

Equation:

$$\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V.$$

- An electron volt is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

Equation:

$$\begin{aligned}
 1 \text{ eV} &= (1.60 \times 10^{-19} \text{ C})(1 \text{ V}) = (1.60 \times 10^{-19} \text{ C})(1 \text{ J/C}) \\
 &= 1.60 \times 10^{-19} \text{ J.}
 \end{aligned}$$

- Mechanical energy is the sum of the kinetic energy and potential energy of a system, that is, $KE + PE$. This sum is a constant.

Conceptual Questions

Exercise:

Problem:

Voltage is the common word for potential difference. Which term is more descriptive, voltage or potential difference?

Exercise:

Problem:

If the voltage between two points is zero, can a test charge be moved between them with zero net work being done? Can this necessarily be done without exerting a force? Explain.

Exercise:

Problem:

What is the relationship between voltage and energy? More precisely, what is the relationship between potential difference and electric potential energy?

Exercise:

Problem: Voltages are always measured between two points. Why?

Exercise:

Problem:

How are units of volts and electron volts related? How do they differ?

Problems & Exercises

Exercise:

Problem:

Find the ratio of speeds of an electron and a negative hydrogen ion (one having an extra electron) accelerated through the same voltage, assuming non-relativistic final speeds. Take the mass of the hydrogen ion to be 1.67×10^{-27} kg.

Solution:

42.8

Exercise:**Problem:**

An evacuated tube uses an accelerating voltage of 40 kV to accelerate electrons to hit a copper plate and produce x rays. Non-relativistically, what would be the maximum speed of these electrons?

Exercise:**Problem:**

A bare helium nucleus has two positive charges and a mass of 6.64×10^{-27} kg. (a) Calculate its kinetic energy in joules at 2.00% of the speed of light. (b) What is this in electron volts? (c) What voltage would be needed to obtain this energy?

Exercise:**Problem: Integrated Concepts**

Singly charged gas ions are accelerated from rest through a voltage of 13.0 V. At what temperature will the average kinetic energy of gas molecules be the same as that given these ions?

Solution:

1.00×10^5 K

Exercise:**Problem: Integrated Concepts**

The temperature near the center of the Sun is thought to be 15 million degrees Celsius (1.5×10^7 °C). Through what voltage must a singly charged ion be accelerated to have the same energy as the average kinetic energy of ions at this temperature?

Exercise:**Problem: Integrated Concepts**

(a) What is the average power output of a heart defibrillator that dissipates 400 J of energy in 10.0 ms? (b) Considering the high-power output, why doesn't the defibrillator produce serious burns?

Solution:

(a) 4×10^4 W

(b) A defibrillator does not cause serious burns because the skin conducts electricity well at high voltages, like those used in defibrillators. The gel used aids in the transfer of energy to the body, and the skin doesn't absorb the energy, but rather lets it pass through to the heart.

Exercise:**Problem: Integrated Concepts**

A lightning bolt strikes a tree, moving 20.0 C of charge through a potential difference of 1.00×10^2 MV. (a) What energy was dissipated? (b) What mass of water could be raised from 15°C to the boiling point and then boiled by this energy? (c) Discuss the damage that could be caused to the tree by the expansion of the boiling steam.

Exercise:

Problem: Integrated Concepts

A 12.0 V battery-operated bottle warmer heats 50.0 g of glass, 2.50×10^2 g of baby formula, and 2.00×10^2 g of aluminum from 20.0°C to 90.0°C. (a) How much charge is moved by the battery? (b) How many electrons per second flow if it takes 5.00 min to warm the formula? (Hint: Assume that the specific heat of baby formula is about the same as the specific heat of water.)

Solution:

(a) 7.40×10^3 C

(b) 1.54×10^{20} electrons per second

Exercise:**Problem: Integrated Concepts**

A battery-operated car utilizes a 12.0 V system. Find the charge the batteries must be able to move in order to accelerate the 750 kg car from rest to 25.0 m/s, make it climb a 2.00×10^2 m high hill, and then cause it to travel at a constant 25.0 m/s by exerting a 5.00×10^2 N force for an hour.

Solution:

3.89×10^6 C

Exercise:**Problem: Integrated Concepts**

Fusion probability is greatly enhanced when appropriate nuclei are brought close together, but mutual Coulomb repulsion must be overcome. This can be done using the kinetic energy of high-temperature gas ions or by accelerating the nuclei toward one another.

(a) Calculate the potential energy of two singly charged nuclei separated by 1.00×10^{-12} m by finding the voltage of one at that distance and multiplying by the charge of the other. (b) At what temperature will atoms of a gas have an average kinetic energy equal to this needed electrical potential energy?

Exercise:

Problem: Unreasonable Results

(a) Find the voltage near a 10.0 cm diameter metal sphere that has 8.00 C of excess positive charge on it. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Solution:

(a) 1.44×10^{12} V

(b) This voltage is very high. A 10.0 cm diameter sphere could never maintain this voltage; it would discharge.

(c) An 8.00 C charge is more charge than can reasonably be accumulated on a sphere of that size.

Exercise:

Problem: Construct Your Own Problem

Consider a battery used to supply energy to a cellular phone. Construct a problem in which you determine the energy that must be supplied by the battery, and then calculate the amount of charge it must be able to move in order to supply this energy. Among the things to be considered are the energy needs and battery voltage. You may need to look ahead to interpret manufacturer's battery ratings in ampere-hours as energy in joules.

Glossary

electric potential

potential energy per unit charge

potential difference (or voltage)

change in potential energy of a charge moved from one point to another, divided by the charge; units of potential difference are joules per coulomb, known as volt

electron volt

the energy given to a fundamental charge accelerated through a potential difference of one volt

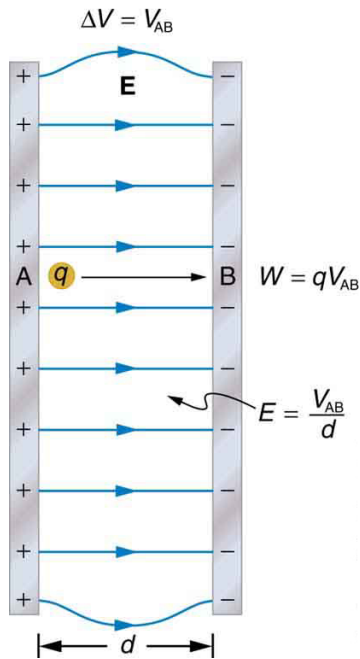
mechanical energy

sum of the kinetic energy and potential energy of a system; this sum is a constant

Electric Potential in a Uniform Electric Field

- Describe the relationship between voltage and electric field.
- Derive an expression for the electric potential and electric field.
- Calculate electric field strength given distance and voltage.

In the previous section, we explored the relationship between voltage and energy. In this section, we will explore the relationship between voltage and electric field. For example, a uniform electric field \mathbf{E} is produced by placing a potential difference (or voltage) ΔV across two parallel metal plates, labeled A and B. (See [\[link\]](#).) Examining this will tell us what voltage is needed to produce a certain electric field strength; it will also reveal a more fundamental relationship between electric potential and electric field. From a physicist's point of view, either ΔV or \mathbf{E} can be used to describe any charge distribution. ΔV is most closely tied to energy, whereas \mathbf{E} is most closely related to force. ΔV is a **scalar** quantity and has no direction, while \mathbf{E} is a **vector** quantity, having both magnitude and direction. (Note that the magnitude of the electric field strength, a scalar quantity, is represented by E below.) The relationship between ΔV and \mathbf{E} is revealed by calculating the work done by the force in moving a charge from point A to point B. But, as noted in [Electric Potential Energy: Potential Difference](#), this is complex for arbitrary charge distributions, requiring calculus. We therefore look at a uniform electric field as an interesting special case.



The relationship between V and E for parallel conducting plates is $E = V/d$. (Note that $\Delta V = V_{AB}$ in magnitude. For a charge that is moved from plate A at higher potential to plate B at lower potential, a minus sign needs to be included as follows:
 $-\Delta V = V_A - V_B = V_{AB}$
 . See the text for details.)

The work done by the electric field in [\[link\]](#) to move a positive charge q from A, the positive plate, higher potential, to B, the negative plate, lower potential, is

Equation:

$$W = -\Delta PE = -q\Delta V.$$

The potential difference between points A and B is

Equation:

$$-\Delta V = -(V_B - V_A) = V_A - V_B = V_{AB}.$$

Entering this into the expression for work yields

Equation:

$$W = qV_{AB}.$$

Work is $W = Fd \cos \theta$; here $\cos \theta = 1$, since the path is parallel to the field, and so $W = Fd$. Since $F = qE$, we see that $W = qEd$. Substituting this expression for work into the previous equation gives

Equation:

$$qEd = qV_{AB}.$$

The charge cancels, and so the voltage between points A and B is seen to be

Equation:

$$\left. \begin{array}{l} V_{AB} = Ed \\ E = \frac{V_{AB}}{d} \end{array} \right\} (\text{uniform } E - \text{field only}),$$

where d is the distance from A to B, or the distance between the plates in [\[link\]](#). Note that the above equation implies the units for electric field are volts per meter. We already know the units for electric field are newtons per coulomb; thus the following relation among units is valid:

Equation:

$$1 \text{ N/C} = 1 \text{ V/m}.$$

Note:

Voltage between Points A and B

Equation:

$$\left. \begin{array}{l} V_{AB} = Ed \\ E = \frac{V_{AB}}{d} \end{array} \right\} (\text{uniform } E - \text{field only}),$$

where d is the distance from A to B, or the distance between the plates.

Example:**What Is the Highest Voltage Possible between Two Plates?**

Dry air will support a maximum electric field strength of about $3.0 \times 10^6 \text{ V/m}$. Above that value, the field creates enough ionization in the air to make the air a conductor. This allows a discharge or spark that reduces the field. What, then, is the maximum voltage between two parallel conducting plates separated by 2.5 cm of dry air?

Strategy

We are given the maximum electric field E between the plates and the distance d between them. The equation $V_{AB} = Ed$ can thus be used to calculate the maximum voltage.

Solution

The potential difference or voltage between the plates is

Equation:

$$V_{AB} = Ed.$$

Entering the given values for E and d gives

Equation:

$$V_{AB} = (3.0 \times 10^6 \text{ V/m})(0.025 \text{ m}) = 7.5 \times 10^4 \text{ V}$$

or

Equation:

$$V_{AB} = 75 \text{ kV}.$$

(The answer is quoted to only two digits, since the maximum field strength is approximate.)

Discussion

One of the implications of this result is that it takes about 75 kV to make a spark jump across a 2.5 cm (1 in.) gap, or 150 kV for a 5 cm spark. This limits the voltages that can exist between conductors, perhaps on a power transmission line. A smaller voltage will cause a spark if there are points on the surface, since points create greater fields than smooth surfaces. Humid air breaks down at a lower field strength, meaning that a smaller voltage will make a spark jump through humid air. The largest voltages can be built up, say with static electricity, on dry days.



A spark chamber is used to trace the paths of high-energy particles. Ionization created by the particles as they pass through the gas between the plates allows a spark to jump. The sparks are

perpendicular to the plates, following electric field lines between them. The potential difference between adjacent plates is not high enough to cause sparks without the ionization produced by particles from accelerator experiments (or cosmic rays).
(credit: Daderot, Wikimedia Commons)

Example:**Field and Force inside an Electron Gun**

(a) An electron gun has parallel plates separated by 4.00 cm and gives electrons 25.0 keV of energy. What is the electric field strength between the plates? (b) What force would this field exert on a piece of plastic with a $0.500\ \mu\text{C}$ charge that gets between the plates?

Strategy

Since the voltage and plate separation are given, the electric field strength can be calculated directly from the expression $E = \frac{V_{AB}}{d}$. Once the electric field strength is known, the force on a charge is found using $\mathbf{F} = q \mathbf{E}$. Since the electric field is in only one direction, we can write this equation in terms of the magnitudes, $F = q E$.

Solution for (a)

The expression for the magnitude of the electric field between two uniform metal plates is

Equation:

$$E = \frac{V_{AB}}{d}.$$

Since the electron is a single charge and is given 25.0 keV of energy, the potential difference must be 25.0 kV. Entering this value for V_{AB} and the plate separation of 0.0400 m, we obtain

Equation:

$$E = \frac{25.0 \text{ kV}}{0.0400 \text{ m}} = 6.25 \times 10^5 \text{ V/m}.$$

Solution for (b)

The magnitude of the force on a charge in an electric field is obtained from the equation

Equation:

$$F = qE.$$

Substituting known values gives

Equation:

$$F = (0.500 \times 10^{-6} \text{ C})(6.25 \times 10^5 \text{ V/m}) = 0.313 \text{ N}.$$

Discussion

Note that the units are newtons, since $1 \text{ V/m} = 1 \text{ N/C}$. The force on the charge is the same no matter where the charge is located between the plates. This is because the electric field is uniform between the plates.

In more general situations, regardless of whether the electric field is uniform, it points in the direction of decreasing potential, because the force on a positive charge is in the direction of \mathbf{E} and also in the direction of lower potential V . Furthermore, the magnitude of \mathbf{E} equals the rate of decrease of V with distance. The faster V decreases over distance, the

greater the electric field. In equation form, the general relationship between voltage and electric field is

Equation:

$$E = - \frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV , takes place. The minus sign tells us that **E** points in the direction of decreasing potential. The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

Note:

Relationship between Voltage and Electric Field

In equation form, the general relationship between voltage and electric field is

Equation:

$$E = - \frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV , takes place. The minus sign tells us that **E** points in the direction of decreasing potential. The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

For continually changing potentials, ΔV and Δs become infinitesimals and differential calculus must be employed to determine the electric field.

Section Summary

- The voltage between points A and B is

Equation:

$$\left. \begin{aligned} V_{AB} &= Ed \\ E &= \frac{V_{AB}}{d} \end{aligned} \right\} (\text{uniform } E - \text{field only}),$$

where d is the distance from A to B, or the distance between the plates.

- In equation form, the general relationship between voltage and electric field is

Equation:

$$E = - \frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV , takes place. The minus sign tells us that \mathbf{E} points in the direction of decreasing potential.) The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

Conceptual Questions

Exercise:

Problem:

Discuss how potential difference and electric field strength are related. Give an example.

Exercise:

Problem:

What is the strength of the electric field in a region where the electric potential is constant?

Exercise:

Problem:

Will a negative charge, initially at rest, move toward higher or lower potential? Explain why.

Problems & Exercises

Exercise:

Problem:

Show that units of V/m and N/C for electric field strength are indeed equivalent.

Exercise:

Problem:

What is the strength of the electric field between two parallel conducting plates separated by 1.00 cm and having a potential difference (voltage) between them of 1.50×10^4 V?

Exercise:

Problem:

The electric field strength between two parallel conducting plates separated by 4.00 cm is 7.50×10^4 V/m. (a) What is the potential difference between the plates? (b) The plate with the lowest potential is taken to be at zero volts. What is the potential 1.00 cm from that plate (and 3.00 cm from the other)?

Solution:

(a) 3.00 kV

(b) 750 V

Exercise:

Problem:

How far apart are two conducting plates that have an electric field strength of 4.50×10^3 V/m between them, if their potential difference is 15.0 kV?

Exercise:

Problem:

(a) Will the electric field strength between two parallel conducting plates exceed the breakdown strength for air ($3.0 \times 10^6 \text{ V/m}$) if the plates are separated by 2.00 mm and a potential difference of $5.0 \times 10^3 \text{ V}$ is applied? (b) How close together can the plates be with this applied voltage?

Solution:

(a) No. The electric field strength between the plates is $2.5 \times 10^6 \text{ V/m}$, which is lower than the breakdown strength for air ($3.0 \times 10^6 \text{ V/m}$).

(b) 1.7 mm

Exercise:**Problem:**

The voltage across a membrane forming a cell wall is 80.0 mV and the membrane is 9.00 nm thick. What is the electric field strength? (The value is surprisingly large, but correct. Membranes are discussed in [Capacitors and Dielectrics](#) and [Nerve Conduction—Electrocardiograms](#).) You may assume a uniform electric field.

Exercise:**Problem:**

Membrane walls of living cells have surprisingly large electric fields across them due to separation of ions. (Membranes are discussed in some detail in [Nerve Conduction—Electrocardiograms](#).) What is the voltage across an 8.00 nm-thick membrane if the electric field strength across it is 5.50 MV/m? You may assume a uniform electric field.

Solution:

44.0 mV

Exercise:**Problem:**

Two parallel conducting plates are separated by 10.0 cm, and one of them is taken to be at zero volts. (a) What is the electric field strength between them, if the potential 8.00 cm from the zero volt plate (and 2.00 cm from the other) is 450 V? (b) What is the voltage between the plates?

Exercise:**Problem:**

Find the maximum potential difference between two parallel conducting plates separated by 0.500 cm of air, given the maximum sustainable electric field strength in air to be 3.0×10^6 V/m.

Solution:

15 kV

Exercise:**Problem:**

A doubly charged ion is accelerated to an energy of 32.0 keV by the electric field between two parallel conducting plates separated by 2.00 cm. What is the electric field strength between the plates?

Exercise:**Problem:**

An electron is to be accelerated in a uniform electric field having a strength of 2.00×10^6 V/m. (a) What energy in keV is given to the electron if it is accelerated through 0.400 m? (b) Over what distance would it have to be accelerated to increase its energy by 50.0 GeV?

Solution:

(a) 800 KeV

(b) 25.0 km

Glossary

scalar

physical quantity with magnitude but no direction

vector

physical quantity with both magnitude and direction

Electrical Potential Due to a Point Charge

- Explain point charges and express the equation for electric potential of a point charge.
- Distinguish between electric potential and electric field.
- Determine the electric potential of a point charge given charge and distance.

Point charges, such as electrons, are among the fundamental building blocks of matter. Furthermore, spherical charge distributions (like on a metal sphere) create external electric fields exactly like a point charge. The electric potential due to a point charge is, thus, a case we need to consider. Using calculus to find the work needed to move a test charge q from a large distance away to a distance of r from a point charge Q , and noting the connection between work and potential ($W = -q\Delta V$), it can be shown that the *electric potential V of a point charge* is

Equation:

$$V = \frac{kQ}{r} \text{ (Point Charge),}$$

where k is a constant equal to $9.0 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$.

Note:

Electric Potential V of a Point Charge

The electric potential V of a point charge is given by

Equation:

$$V = \frac{kQ}{r} \text{ (Point Charge).}$$

The potential at infinity is chosen to be zero. Thus V for a point charge decreases with distance, whereas \mathbf{E} for a point charge decreases with distance squared:

Equation:

$$E = \frac{F}{q} = \frac{kQ}{r^2}.$$

Recall that the electric potential V is a scalar and has no direction, whereas the electric field \mathbf{E} is a vector. To find the voltage due to a combination of point charges, you add the individual voltages as numbers. To find the total electric field, you must add the individual fields as *vectors*, taking magnitude and direction into account. This is consistent with the fact that V is closely associated with energy, a scalar, whereas \mathbf{E} is closely associated with force, a vector.

Example:

What Voltage Is Produced by a Small Charge on a Metal Sphere?

Charges in static electricity are typically in the nanocoulomb (nC) to microcoulomb (μC) range. What is the voltage 5.00 cm away from the center of a 1-cm diameter metal sphere that has a -3.00 nC static charge?

Strategy

As we have discussed in [Electric Charge and Electric Field](#), charge on a metal sphere spreads out uniformly and produces a field like that of a point charge located at its center. Thus we can find the voltage using the equation $V = kQ/r$.

Solution

Entering known values into the expression for the potential of a point charge, we obtain

Equation:

$$\begin{aligned}
 V &= k \frac{Q}{r} \\
 &= (8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \left(\frac{-3.00 \times 10^{-9} \text{ C}}{5.00 \times 10^{-2} \text{ m}} \right) \\
 &= -539 \text{ V}.
 \end{aligned}$$

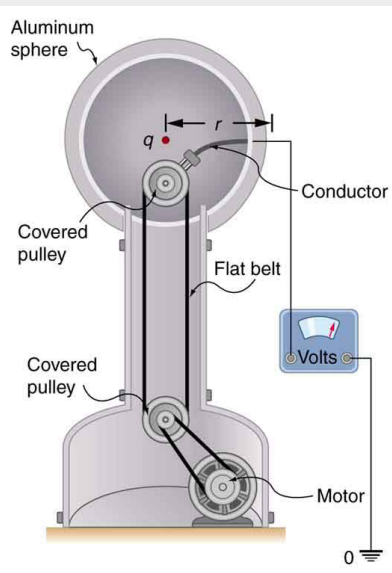
Discussion

The negative value for voltage means a positive charge would be attracted from a larger distance, since the potential is lower (more negative) than at larger distances. Conversely, a negative charge would be repelled, as expected.

Example:

What Is the Excess Charge on a Van de Graaff Generator

A demonstration Van de Graaff generator has a 25.0 cm diameter metal sphere that produces a voltage of 100 kV near its surface. (See [\[link\]](#).) What excess charge resides on the sphere? (Assume that each numerical value here is shown with three significant figures.)



The voltage of this demonstration Van de Graaff generator is measured

between the charged sphere and ground. Earth's potential is taken to be zero as a reference. The potential of the charged conducting sphere is the same as that of an equal point charge at its center.

Strategy

The potential on the surface will be the same as that of a point charge at the center of the sphere, 12.5 cm away. (The radius of the sphere is 12.5 cm.)

We can thus determine the excess charge using the equation

Equation:

$$V = \frac{kQ}{r}.$$

Solution

Solving for Q and entering known values gives

Equation:

$$\begin{aligned} Q &= \frac{rV}{k} \\ &= \frac{(0.125 \text{ m})(100 \times 10^3 \text{ V})}{8.99 \times 10^9 \text{ N} \cdot \text{m}^2 / \text{C}^2} \\ &= 1.39 \times 10^{-6} \text{ C} = 1.39 \text{ } \mu\text{C}. \end{aligned}$$

Discussion

This is a relatively small charge, but it produces a rather large voltage. We have another indication here that it is difficult to store isolated charges.

The voltages in both of these examples could be measured with a meter that compares the measured potential with ground potential. Ground potential is often taken to be zero (instead of taking the potential at infinity to be zero). It is the potential difference between two points that is of importance, and very often there is a tacit assumption that some reference point, such as Earth or a very distant point, is at zero potential. As noted in [Electric Potential Energy: Potential Difference](#), this is analogous to taking sea level as $h = 0$ when considering gravitational potential energy, $PE_g = mgh$.

Section Summary

- Electric potential of a point charge is $V = kQ/r$.
- Electric potential is a scalar, and electric field is a vector. Addition of voltages as numbers gives the voltage due to a combination of point charges, whereas addition of individual fields as vectors gives the total electric field.

Conceptual Questions

Exercise:

Problem:

In what region of space is the potential due to a uniformly charged sphere the same as that of a point charge? In what region does it differ from that of a point charge?

Exercise:

Problem:

Can the potential of a non-uniformly charged sphere be the same as that of a point charge? Explain.

Problems & Exercises

Exercise:

Problem:

A 0.500 cm diameter plastic sphere, used in a static electricity demonstration, has a uniformly distributed 40.0 pC charge on its surface. What is the potential near its surface?

Solution:

144 V

Exercise:**Problem:**

What is the potential 0.530×10^{-10} m from a proton (the average distance between the proton and electron in a hydrogen atom)?

Exercise:**Problem:**

(a) A sphere has a surface uniformly charged with 1.00 C. At what distance from its center is the potential 5.00 MV? (b) What does your answer imply about the practical aspect of isolating such a large charge?

Solution:

(a) 1.80 km

(b) A charge of 1 C is a very large amount of charge; a sphere of radius 1.80 km is not practical.

Exercise:**Problem:**

How far from a 1.00 μC point charge will the potential be 100 V? At what distance will it be 2.00×10^2 V?

Exercise:

Problem:

What are the sign and magnitude of a point charge that produces a potential of -2.00 V at a distance of 1.00 mm ?

Solution:

$$-2.22 \times 10^{-13}\text{ C}$$

Exercise:**Problem:**

If the potential due to a point charge is $5.00 \times 10^2\text{ V}$ at a distance of 15.0 m , what are the sign and magnitude of the charge?

Exercise:**Problem:**

In nuclear fission, a nucleus splits roughly in half. (a) What is the potential $2.00 \times 10^{-14}\text{ m}$ from a fragment that has 46 protons in it? (b) What is the potential energy in MeV of a similarly charged fragment at this distance?

Solution:

(a) $3.31 \times 10^6\text{ V}$

(b) 152 MeV

Exercise:**Problem:**

A research Van de Graaff generator has a 2.00-m -diameter metal sphere with a charge of 5.00 mC on it. (a) What is the potential near its surface? (b) At what distance from its center is the potential 1.00 MV ? (c) An oxygen atom with three missing electrons is released near the Van de Graaff generator. What is its energy in MeV at this distance?

Exercise:**Problem:**

An electrostatic paint sprayer has a 0.200-m-diameter metal sphere at a potential of 25.0 kV that repels paint droplets onto a grounded object. (a) What charge is on the sphere? (b) What charge must a 0.100-mg drop of paint have to arrive at the object with a speed of 10.0 m/s?

Solution:

(a) $2.78 \times 10^{-7} \text{ C}$

(b) $2.00 \times 10^{-10} \text{ C}$

Exercise:**Problem:**

In one of the classic nuclear physics experiments at the beginning of the 20th century, an alpha particle was accelerated toward a gold nucleus, and its path was substantially deflected by the Coulomb interaction. If the energy of the doubly charged alpha nucleus was 5.00 MeV, how close to the gold nucleus (79 protons) could it come before being deflected?

Exercise:**Problem:**

(a) What is the potential between two points situated 10 cm and 20 cm from a $3.0 \mu\text{C}$ point charge? (b) To what location should the point at 20 cm be moved to increase this potential difference by a factor of two?

Exercise:**Problem: Unreasonable Results**

(a) What is the final speed of an electron accelerated from rest through a voltage of 25.0 MV by a negatively charged Van de Graaff terminal?

(b) What is unreasonable about this result?

(c) Which assumptions are responsible?

Solution:

(a) $2.96 \times 10^9 \text{ m/s}$

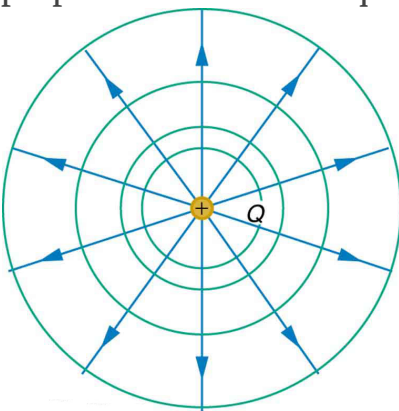
(b) This velocity is far too great. It is faster than the speed of light.

(c) The assumption that the speed of the electron is far less than that of light and that the problem does not require a relativistic treatment produces an answer greater than the speed of light.

Equipotential Lines

- Explain equipotential lines and equipotential surfaces.
- Describe the action of grounding an electrical appliance.
- Compare electric field and equipotential lines.

We can represent electric potentials (voltages) pictorially, just as we drew pictures to illustrate electric fields. Of course, the two are related. Consider [\[link\]](#), which shows an isolated positive point charge and its electric field lines. Electric field lines radiate out from a positive charge and terminate on negative charges. While we use blue arrows to represent the magnitude and direction of the electric field, we use green lines to represent places where the electric potential is constant. These are called **equipotential lines** in two dimensions, or *equipotential surfaces* in three dimensions. The term *equipotential* is also used as a noun, referring to an equipotential line or surface. The potential for a point charge is the same anywhere on an imaginary sphere of radius r surrounding the charge. This is true since the potential for a point charge is given by $V = kQ/r$ and, thus, has the same value at any point that is a given distance r from the charge. An equipotential sphere is a circle in the two-dimensional view of [\[link\]](#). Since the electric field lines point radially away from the charge, they are perpendicular to the equipotential lines.



An isolated point charge Q with its electric field lines in blue and equipotential lines

in green. The potential is the same along each equipotential line, meaning that no work is required to move a charge anywhere along one of those lines. Work is needed to move a charge from one equipotential line to another. Equipotential lines are perpendicular to electric field lines in every case.

It is important to note that *equipotential lines are always perpendicular to electric field lines*. No work is required to move a charge along an equipotential, since $\Delta V = 0$. Thus the work is

Equation:

$$W = -\Delta \text{PE} = -q\Delta V = 0.$$

Work is zero if force is perpendicular to motion. Force is in the same direction as \mathbf{E} , so that motion along an equipotential must be perpendicular to \mathbf{E} . More precisely, work is related to the electric field by

Equation:

$$W = Fd \cos \theta = qEd \cos \theta = 0.$$

Note that in the above equation, E and F symbolize the magnitudes of the electric field strength and force, respectively. Neither q nor \mathbf{E} nor d is zero, and so $\cos \theta$ must be 0, meaning θ must be 90° . In other words, motion along an equipotential is perpendicular to \mathbf{E} .

One of the rules for static electric fields and conductors is that the electric field must be perpendicular to the surface of any conductor. This implies that a *conductor is an equipotential surface in static situations*. There can be no voltage difference across the surface of a conductor, or charges will flow. One of the uses of this fact is that a conductor can be fixed at zero volts by connecting it to the earth with a good conductor—a process called **grounding**. Grounding can be a useful safety tool. For example, grounding the metal case of an electrical appliance ensures that it is at zero volts relative to the earth.

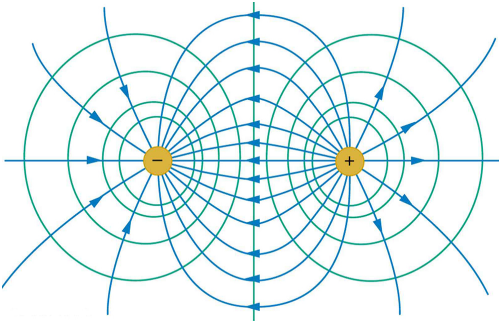
Note:

Grounding

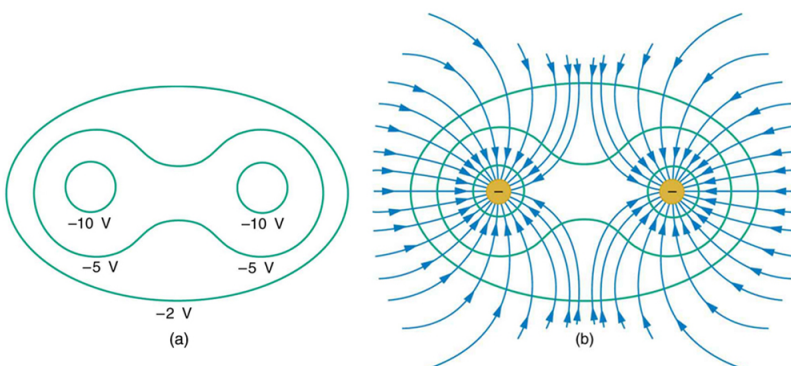
A conductor can be fixed at zero volts by connecting it to the earth with a good conductor—a process called grounding.

Because a conductor is an equipotential, it can replace any equipotential surface. For example, in [\[link\]](#) a charged spherical conductor can replace the point charge, and the electric field and potential surfaces outside of it will be unchanged, confirming the contention that a spherical charge distribution is equivalent to a point charge at its center.

[\[link\]](#) shows the electric field and equipotential lines for two equal and opposite charges. Given the electric field lines, the equipotential lines can be drawn simply by making them perpendicular to the electric field lines. Conversely, given the equipotential lines, as in [\[link\]\(a\)](#), the electric field lines can be drawn by making them perpendicular to the equipotentials, as in [\[link\]\(b\)](#).



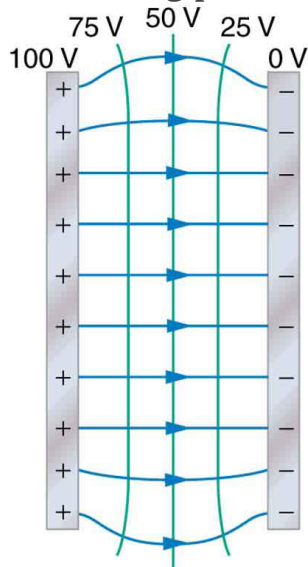
The electric field lines and equipotential lines for two equal but opposite charges. The equipotential lines can be drawn by making them perpendicular to the electric field lines, if those are known. Note that the potential is greatest (most positive) near the positive charge and least (most negative) near the negative charge.



(a) These equipotential lines might be measured with a voltmeter in a laboratory experiment. (b) The corresponding electric field lines are found by drawing them

perpendicular to the equipotentials. Note that these fields are consistent with two equal negative charges.

One of the most important cases is that of the familiar parallel conducting plates shown in [\[link\]](#). Between the plates, the equipotentials are evenly spaced and parallel. The same field could be maintained by placing conducting plates at the equipotential lines at the potentials shown.



The electric field and equipotential lines between two metal plates.

An important application of electric fields and equipotential lines involves the heart. The heart relies on electrical signals to maintain its rhythm. The movement of electrical signals causes the chambers of the heart to contract and relax. When a person has a heart attack, the movement of these electrical signals may be disturbed. An artificial pacemaker and a

defibrillator can be used to initiate the rhythm of electrical signals. The equipotential lines around the heart, the thoracic region, and the axis of the heart are useful ways of monitoring the structure and functions of the heart. An electrocardiogram (ECG) measures the small electric signals being generated during the activity of the heart. More about the relationship between electric fields and the heart is discussed in [Energy Stored in Capacitors](#).

Note:**PhET Explorations: Charges and Fields**

Move point charges around on the playing field and then view the electric field, voltages, equipotential lines, and more. It's colorful, it's dynamic, it's free.

https://phet.colorado.edu/sims/html/charges-and-fields/latest/charges-and-fields_en.html

Section Summary

- An equipotential line is a line along which the electric potential is constant.
- An equipotential surface is a three-dimensional version of equipotential lines.
- Equipotential lines are always perpendicular to electric field lines.
- The process by which a conductor can be fixed at zero volts by connecting it to the earth with a good conductor is called grounding.

Conceptual Questions

Exercise:**Problem:**

What is an equipotential line? What is an equipotential surface?

Exercise:

Problem:

Explain in your own words why equipotential lines and surfaces must be perpendicular to electric field lines.

Exercise:

Problem: Can different equipotential lines cross? Explain.

Problems & Exercises

Exercise:

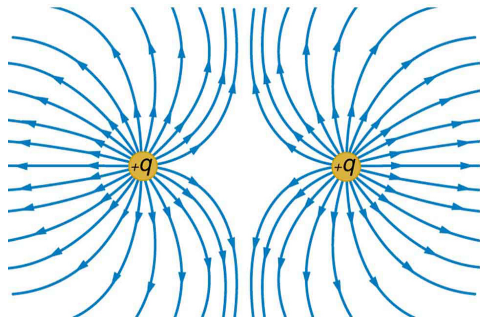
Problem:

(a) Sketch the equipotential lines near a point charge $+q$. Indicate the direction of increasing potential. (b) Do the same for a point charge $-3q$.

Exercise:

Problem:

Sketch the equipotential lines for the two equal positive charges shown in [\[link\]](#). Indicate the direction of increasing potential.



The electric field near two equal positive charges is directed away from each of the charges.

Exercise:

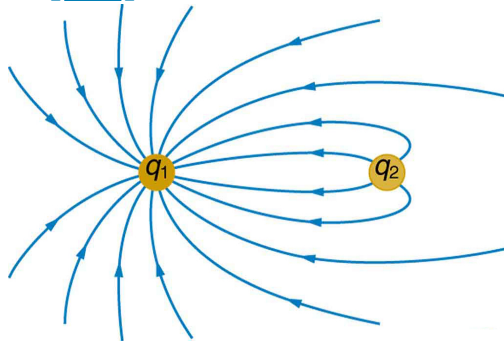
Problem:

[\[link\]](#) shows the electric field lines near two charges q_1 and q_2 , the first having a magnitude four times that of the second. Sketch the equipotential lines for these two charges, and indicate the direction of increasing potential.

Exercise:

Problem:

Sketch the equipotential lines a long distance from the charges shown in [\[link\]](#). Indicate the direction of increasing potential.



The electric field near
two charges.

Exercise:

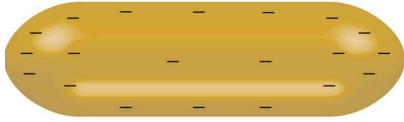
Problem:

Sketch the equipotential lines in the vicinity of two opposite charges, where the negative charge is three times as great in magnitude as the positive. See [\[link\]](#) for a similar situation. Indicate the direction of increasing potential.

Exercise:

Problem:

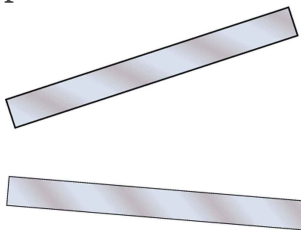
Sketch the equipotential lines in the vicinity of the negatively charged conductor in [\[link\]](#). How will these equipotentials look a long distance from the object?



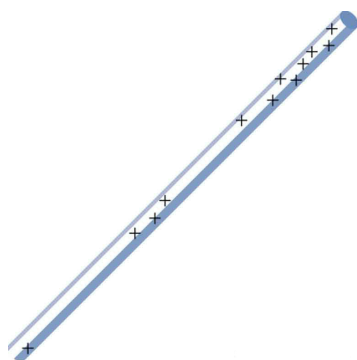
A negatively charged conductor.

Exercise:**Problem:**

Sketch the equipotential lines surrounding the two conducting plates shown in [\[link\]](#), given the top plate is positive and the bottom plate has an equal amount of negative charge. Be certain to indicate the distribution of charge on the plates. Is the field strongest where the plates are closest? Why should it be?

**Exercise:****Problem:**

(a) Sketch the electric field lines in the vicinity of the charged insulator in [\[link\]](#). Note its non-uniform charge distribution. (b) Sketch equipotential lines surrounding the insulator. Indicate the direction of increasing potential.



A charged insulating rod such as might be used in a classroom demonstration.

Exercise:

Problem:

The naturally occurring charge on the ground on a fine day out in the open country is -1.00 nC/m^2 . (a) What is the electric field relative to ground at a height of 3.00 m? (b) Calculate the electric potential at this height. (c) Sketch electric field and equipotential lines for this scenario.

Exercise:

Problem:

The lesser electric ray (*Narcine bancroftii*) maintains an incredible charge on its head and a charge equal in magnitude but opposite in sign on its tail ([\[link\]](#)). (a) Sketch the equipotential lines surrounding the ray. (b) Sketch the equipotentials when the ray is near a ship with a conducting surface. (c) How could this charge distribution be of use to the ray?



Lesser electric ray (*Narcine bancroftii*) (credit: National Oceanic and Atmospheric Administration, NOAA's Fisheries Collection).

Glossary

equipotential line

a line along which the electric potential is constant

grounding

fixing a conductor at zero volts by connecting it to the earth or ground

Introduction to Electric Current, Resistance, and Ohm's Law

class="introduction"

Electric energy in massive quantities is transmitted from this hydroelectric facility, the Srisailem power station located along the Krishna River in India, by the movement of charge—that is, by electric current.
(credit: Chintohere, Wikimedia Commons)



The flicker of numbers on a handheld calculator, nerve impulses carrying signals of vision to the brain, an ultrasound device sending a signal to a computer screen, the brain sending a message for a baby to twitch its toes, an electric train pulling its load over a mountain pass, a hydroelectric plant sending energy to metropolitan and rural users—these and many other examples of electricity involve *electric current, the movement of charge*. Humankind has indeed harnessed electricity, the basis of technology, to improve our quality of life. Whereas the previous two chapters concentrated on static electricity and the fundamental force underlying its behavior, the next few chapters will be devoted to electric and magnetic phenomena involving current. In addition to exploring applications of electricity, we shall gain new insights into nature—in particular, the fact that all magnetism results from electric current.

Current

- Define electric current, ampere, and drift velocity
- Describe the direction of charge flow in conventional current.
- Use drift velocity to calculate current and vice versa.

Electric Current

Electric current is defined to be the rate at which charge flows. A large current, such as that used to start a truck engine, moves a large amount of charge in a small time, whereas a small current, such as that used to operate a hand-held calculator, moves a small amount of charge over a long period of time. In equation form, **electric current** I is defined to be

Equation:

$$I = \frac{\Delta Q}{\Delta t},$$

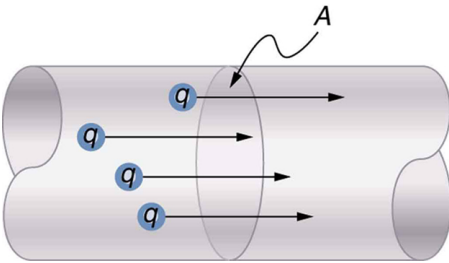
where ΔQ is the amount of charge passing through a given area in time Δt . (As in previous chapters, initial time is often taken to be zero, in which case $\Delta t = t$.) (See [\[link\]](#).) The SI unit for current is the **ampere** (A), named for the French physicist André-Marie Ampère (1775–1836). Since $I = \Delta Q / \Delta t$, we see that an ampere is one coulomb per second:

Equation:

$$1 \text{ A} = 1 \text{ C/s}$$

Not only are fuses and circuit breakers rated in amperes (or amps), so are many electrical appliances.

Current = flow of charge



The rate of flow of charge is current. An ampere is the flow of one coulomb through an area in one second.

Example:

Calculating Currents: Current in a Truck Battery and a Handheld Calculator

(a) What is the current involved when a truck battery sets in motion 720 C of charge in 4.00 s while starting an engine? (b) How long does it take 1.00 C of charge to flow through a handheld calculator if a 0.300-mA current is flowing?

Strategy

We can use the definition of current in the equation $I = \Delta Q / \Delta t$ to find the current in part (a), since charge and time are given. In part (b), we rearrange the definition of current and use the given values of charge and current to find the time required.

Solution for (a)

Entering the given values for charge and time into the definition of current gives

Equation:

$$\begin{aligned} I &= \frac{\Delta Q}{\Delta t} = \frac{720 \text{ C}}{4.00 \text{ s}} = 180 \text{ C/s} \\ &= 180 \text{ A.} \end{aligned}$$

Discussion for (a)

This large value for current illustrates the fact that a large charge is moved in a small amount of time. The currents in these “starter motors” are fairly large because large frictional forces need to be overcome when setting something in motion.

Solution for (b)

Solving the relationship $I = \Delta Q / \Delta t$ for time Δt , and entering the known values for charge and current gives

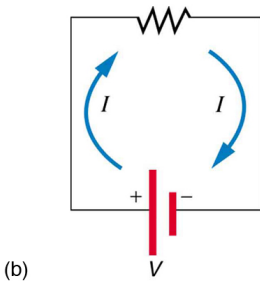
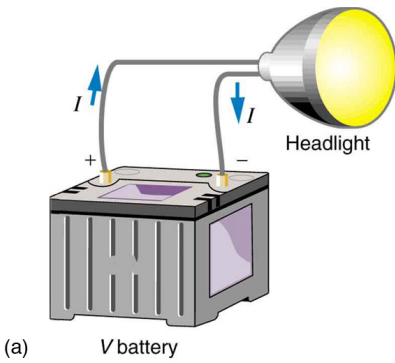
Equation:

$$\begin{aligned}\Delta t &= \frac{\Delta Q}{I} = \frac{1.00 \text{ C}}{0.300 \times 10^{-3} \text{ C/s}} \\ &= 3.33 \times 10^3 \text{ s.}\end{aligned}$$

Discussion for (b)

This time is slightly less than an hour. The small current used by the handheld calculator takes a much longer time to move a smaller charge than the large current of the truck starter. So why can we operate our calculators only seconds after turning them on? It's because calculators require very little energy. Such small current and energy demands allow handheld calculators to operate from solar cells or to get many hours of use out of small batteries. Remember, calculators do not have moving parts in the same way that a truck engine has with cylinders and pistons, so the technology requires smaller currents.

[\[link\]](#) shows a simple circuit and the standard schematic representation of a battery, conducting path, and load (a resistor). Schematics are very useful in visualizing the main features of a circuit. A single schematic can represent a wide variety of situations. The schematic in [\[link\]](#) (b), for example, can represent anything from a truck battery connected to a headlight lighting the street in front of the truck to a small battery connected to a penlight lighting a keyhole in a door. Such schematics are useful because the analysis is the same for a wide variety of situations. We need to understand a few schematics to apply the concepts and analysis to many more situations.



(a) A simple electric circuit. A closed path for current to flow through is supplied by conducting wires connecting a load to the terminals of a battery. (b) In this schematic, the battery is represented by the two parallel red lines, conducting wires are shown as straight lines, and the zigzag represents the load. The schematic represents a wide

variety of similar
circuits.

Note that the direction of current flow in [\[link\]](#) is from positive to negative. *The direction of conventional current is the direction that positive charge would flow.* Depending on the situation, positive charges, negative charges, or both may move. In metal wires, for example, current is carried by electrons—that is, negative charges move. In ionic solutions, such as salt water, both positive and negative charges move. This is also true in nerve cells. A Van de Graaff generator used for nuclear research can produce a current of pure positive charges, such as protons. [\[link\]](#) illustrates the movement of charged particles that compose a current. The fact that conventional current is taken to be in the direction that positive charge would flow can be traced back to American politician and scientist Benjamin Franklin in the 1700s. He named the type of charge associated with electrons negative, long before they were known to carry current in so many situations. Franklin, in fact, was totally unaware of the small-scale structure of electricity.

It is important to realize that there is an electric field in conductors responsible for producing the current, as illustrated in [\[link\]](#). Unlike static electricity, where a conductor in equilibrium cannot have an electric field in it, conductors carrying a current have an electric field and are not in static equilibrium. An electric field is needed to supply energy to move the charges.

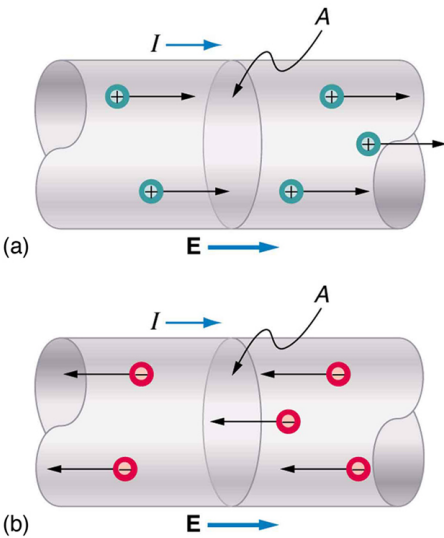
Note:

Making Connections: Take-Home Investigation—Electric Current Illustration

Find a straw and little peas that can move freely in the straw. Place the straw flat on a table and fill the straw with peas. When you pop one pea in at one end, a different pea should pop out the other end. This demonstration is an analogy for an electric current. Identify what compares

to the electrons and what compares to the supply of energy. What other analogies can you find for an electric current?

Note that the flow of peas is based on the peas physically bumping into each other; electrons flow due to mutually repulsive electrostatic forces.



Current I is the rate at which charge moves through an area A , such as the cross-section of a wire.

Conventional current is defined to move in the direction of the electric field. (a)

Positive charges move in the direction of the electric field and the same direction as conventional current.

(b) Negative charges move in the direction opposite to the electric field. Conventional

current is in the direction opposite to the movement of negative charge. The flow of electrons is sometimes referred to as electronic flow.

Example:**Calculating the Number of Electrons that Move through a Calculator**

If the 0.300-mA current through the calculator mentioned in the [\[link\]](#) example is carried by electrons, how many electrons per second pass through it?

Strategy

The current calculated in the previous example was defined for the flow of positive charge. For electrons, the magnitude is the same, but the sign is opposite, $I_{\text{electrons}} = -0.300 \times 10^{-3} \text{ C/s}$. Since each electron (e^-) has a charge of $-1.60 \times 10^{-19} \text{ C}$, we can convert the current in coulombs per second to electrons per second.

Solution

Starting with the definition of current, we have

Equation:

$$I_{\text{electrons}} = \frac{\Delta Q_{\text{electrons}}}{\Delta t} = \frac{-0.300 \times 10^{-3} \text{ C}}{\text{s}}.$$

We divide this by the charge per electron, so that

Equation:

$$\begin{aligned} \frac{e^-}{\text{s}} &= \frac{-0.300 \times 10^{-3} \text{ C}}{\text{s}} \times \frac{1 e^-}{-1.60 \times 10^{-19} \text{ C}} \\ &= 1.88 \times 10^{15} \frac{e^-}{\text{s}}. \end{aligned}$$

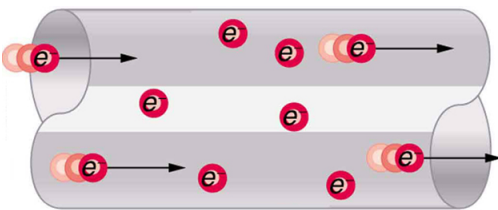
Discussion

There are so many charged particles moving, even in small currents, that individual charges are not noticed, just as individual water molecules are not noticed in water flow. Even more amazing is that they do not always keep moving forward like soldiers in a parade. Rather they are like a crowd of people with movement in different directions but a general trend to move forward. There are lots of collisions with atoms in the metal wire and, of course, with other electrons.

Drift Velocity

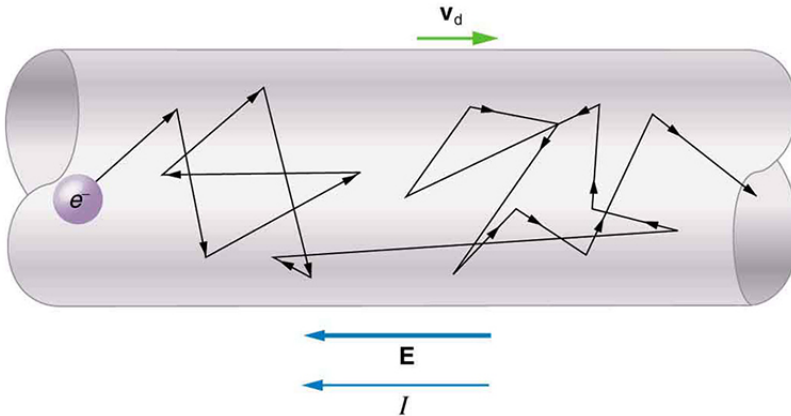
Electrical signals are known to move very rapidly. Telephone conversations carried by currents in wires cover large distances without noticeable delays. Lights come on as soon as a switch is flicked. Most electrical signals carried by currents travel at speeds on the order of 10^8 m/s, a significant fraction of the speed of light. Interestingly, the individual charges that make up the current move *much* more slowly on average, typically drifting at speeds on the order of 10^{-4} m/s. How do we reconcile these two speeds, and what does it tell us about standard conductors?

The high speed of electrical signals results from the fact that the force between charges acts rapidly at a distance. Thus, when a free charge is forced into a wire, as in [\[link\]](#), the incoming charge pushes other charges ahead of it, which in turn push on charges farther down the line. The density of charge in a system cannot easily be increased, and so the signal is passed on rapidly. The resulting electrical shock wave moves through the system at nearly the speed of light. To be precise, this rapidly moving signal or shock wave is a rapidly propagating change in electric field.



When charged particles are forced into this volume of a conductor, an equal number are quickly forced to leave. The repulsion between like charges makes it difficult to increase the number of charges in a volume. Thus, as one charge enters, another leaves almost immediately, carrying the signal rapidly forward.

Good conductors have large numbers of free charges in them. In metals, the free charges are free electrons. [\[link\]](#) shows how free electrons move through an ordinary conductor. The distance that an individual electron can move between collisions with atoms or other electrons is quite small. The electron paths thus appear nearly random, like the motion of atoms in a gas. But there is an electric field in the conductor that causes the electrons to drift in the direction shown (opposite to the field, since they are negative). The **drift velocity** v_d is the average velocity of the free charges. Drift velocity is quite small, since there are so many free charges. If we have an estimate of the density of free electrons in a conductor, we can calculate the drift velocity for a given current. The larger the density, the lower the velocity required for a given current.



Free electrons moving in a conductor make many collisions with other electrons and atoms. The path of one electron is shown. The average velocity of the free charges is called the drift velocity, v_d , and it is in the direction opposite to the electric field for electrons. The collisions normally transfer energy to the conductor, requiring a constant supply of energy to maintain a steady current.

Note:

Conduction of Electricity and Heat

Good electrical conductors are often good heat conductors, too. This is because large numbers of free electrons can carry electrical current and can transport thermal energy.

The free-electron collisions transfer energy to the atoms of the conductor. The electric field does work in moving the electrons through a distance, but that work does not increase the kinetic energy (nor speed, therefore) of the electrons. The work is transferred to the conductor's atoms, possibly

increasing temperature. Thus a continuous power input is required to keep a current flowing. An exception, of course, is found in superconductors, for reasons we shall explore in a later chapter. Superconductors can have a steady current without a continual supply of energy—a great energy savings. In contrast, the supply of energy can be useful, such as in a lightbulb filament. The supply of energy is necessary to increase the temperature of the tungsten filament, so that the filament glows.

Note:

Making Connections: Take-Home Investigation—Filament Observations

Find a lightbulb with a filament. Look carefully at the filament and describe its structure. To what points is the filament connected?

We can obtain an expression for the relationship between current and drift velocity by considering the number of free charges in a segment of wire, as illustrated in [\[link\]](#). The number of free charges per unit volume is given the symbol n and depends on the material. The shaded segment has a volume Ax , so that the number of free charges in it is nAx . The charge ΔQ in this segment is thus $qnAx$, where q is the amount of charge on each carrier. (Recall that for electrons, q is -1.60×10^{-19} C.) Current is charge moved per unit time; thus, if all the original charges move out of this segment in time Δt , the current is

Equation:

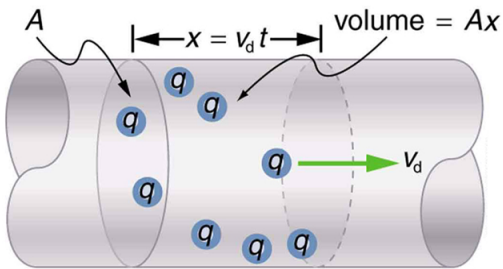
$$I = \frac{\Delta Q}{\Delta t} = \frac{qnAx}{\Delta t}.$$

Note that $x/\Delta t$ is the magnitude of the drift velocity, v_d , since the charges move an average distance x in a time Δt . Rearranging terms gives

Equation:

$$I = nqAv_d,$$

where I is the current through a wire of cross-sectional area A made of a material with a free charge density n . The carriers of the current each have charge q and move with a drift velocity of magnitude v_d .



All the charges in the shaded volume of this wire move out in a time t , having a drift velocity of magnitude $v_d = x/t$. See text for further discussion.

Note that simple drift velocity is not the entire story. The speed of an electron is much greater than its drift velocity. In addition, not all of the electrons in a conductor can move freely, and those that do might move somewhat faster or slower than the drift velocity. So what do we mean by free electrons? Atoms in a metallic conductor are packed in the form of a lattice structure. Some electrons are far enough away from the atomic nuclei that they do not experience the attraction of the nuclei as much as the inner electrons do. These are the free electrons. They are not bound to a single atom but can instead move freely among the atoms in a “sea” of electrons. These free electrons respond by accelerating when an electric field is applied. Of course as they move they collide with the atoms in the lattice and other electrons, generating thermal energy, and the conductor gets warmer. In an insulator, the organization of the atoms and the structure do not allow for such free electrons.

Example:**Calculating Drift Velocity in a Common Wire**

Calculate the drift velocity of electrons in a 12-gauge copper wire (which has a diameter of 2.053 mm) carrying a 20.0-A current, given that there is one free electron per copper atom. (Household wiring often contains 12-gauge copper wire, and the maximum current allowed in such wire is usually 20 A.) The density of copper is $8.80 \times 10^3 \text{ kg/m}^3$.

Strategy

We can calculate the drift velocity using the equation $I = nqAv_d$. The current $I = 20.0 \text{ A}$ is given, and $q = -1.60 \times 10^{-19} \text{ C}$ is the charge of an electron. We can calculate the area of a cross-section of the wire using the formula $A = \pi r^2$, where r is one-half the given diameter, 2.053 mm. We are given the density of copper, $8.80 \times 10^3 \text{ kg/m}^3$, and the periodic table shows that the atomic mass of copper is 63.54 g/mol. We can use these two quantities along with Avogadro's number, $6.02 \times 10^{23} \text{ atoms/mol}$, to determine n , the number of free electrons per cubic meter.

Solution

First, calculate the density of free electrons in copper. There is one free electron per copper atom. Therefore, is the same as the number of copper atoms per m^3 . We can now find n as follows:

Equation:

$$\begin{aligned} n &= \frac{1 e^-}{\text{atom}} \times \frac{6.02 \times 10^{23} \text{ atoms}}{\text{mol}} \times \frac{1 \text{ mol}}{63.54 \text{ g}} \times \frac{1000 \text{ g}}{\text{kg}} \times \frac{8.80 \times 10^3 \text{ kg}}{1 \text{ m}^3} \\ &= 8.342 \times 10^{28} e^-/\text{m}^3. \end{aligned}$$

The cross-sectional area of the wire is

Equation:

$$\begin{aligned} A &= \pi r^2 \\ &= \pi \left(\frac{2.053 \times 10^{-3} \text{ m}}{2} \right)^2 \\ &= 3.310 \times 10^{-6} \text{ m}^2. \end{aligned}$$

Rearranging $I = nqAv_d$ to isolate drift velocity gives

Equation:

$$\begin{aligned}
 v_d &= \frac{I}{nqA} \\
 &= \frac{20.0 \text{ A}}{(8.342 \times 10^{28} / \text{m}^3)(-1.60 \times 10^{-19} \text{ C})(3.310 \times 10^{-6} \text{ m}^2)} \\
 &= -4.53 \times 10^{-4} \text{ m/s}.
 \end{aligned}$$

Discussion

The minus sign indicates that the negative charges are moving in the direction opposite to conventional current. The small value for drift velocity (on the order of 10^{-4} m/s) confirms that the signal moves on the order of 10^{12} times faster (about 10^8 m/s) than the charges that carry it.

Section Summary

- Electric current I is the rate at which charge flows, given by

Equation:

$$I = \frac{\Delta Q}{\Delta t},$$

where ΔQ is the amount of charge passing through an area in time Δt .

- The direction of conventional current is taken as the direction in which positive charge moves.
- The SI unit for current is the ampere (A), where $1 \text{ A} = 1 \text{ C/s}$.
- Current is the flow of free charges, such as electrons and ions.
- Drift velocity v_d is the average speed at which these charges move.
- Current I is proportional to drift velocity v_d , as expressed in the relationship $I = nqAv_d$. Here, I is the current through a wire of cross-sectional area A . The wire's material has a free-charge density n , and each carrier has charge q and a drift velocity v_d .
- Electrical signals travel at speeds about 10^{12} times greater than the drift velocity of free electrons.

Conceptual Questions

Exercise:

Problem:

Can a wire carry a current and still be neutral—that is, have a total charge of zero? Explain.

Exercise:**Problem:**

Car batteries are rated in ampere-hours ($A \cdot h$). To what physical quantity do ampere-hours correspond (voltage, charge, . . .), and what relationship do ampere-hours have to energy content?

Exercise:**Problem:**

If two different wires having identical cross-sectional areas carry the same current, will the drift velocity be higher or lower in the better conductor? Explain in terms of the equation $v_d = \frac{I}{nqA}$, by considering how the density of charge carriers n relates to whether or not a material is a good conductor.

Exercise:**Problem:**

Why are two conducting paths from a voltage source to an electrical device needed to operate the device?

Exercise:**Problem:**

In cars, one battery terminal is connected to the metal body. How does this allow a single wire to supply current to electrical devices rather than two wires?

Exercise:

Problem:

Why isn't a bird sitting on a high-voltage power line electrocuted? Contrast this with the situation in which a large bird hits two wires simultaneously with its wings.

Problems & Exercises**Exercise:****Problem:**

What is the current in milliamperes produced by the solar cells of a pocket calculator through which 4.00 C of charge passes in 4.00 h?

Solution:

0.278 mA

Exercise:**Problem:**

A total of 600 C of charge passes through a flashlight in 0.500 h. What is the average current?

Exercise:**Problem:**

What is the current when a typical static charge of $0.250\ \mu\text{C}$ moves from your finger to a metal doorknob in $1.00\ \mu\text{s}$?

Solution:

0.250 A

Exercise:

Problem:

Find the current when 2.00 nC jumps between your comb and hair over a 0.500 - μ s time interval.

Exercise:**Problem:**

A large lightning bolt had a 20,000-A current and moved 30.0 C of charge. What was its duration?

Solution:

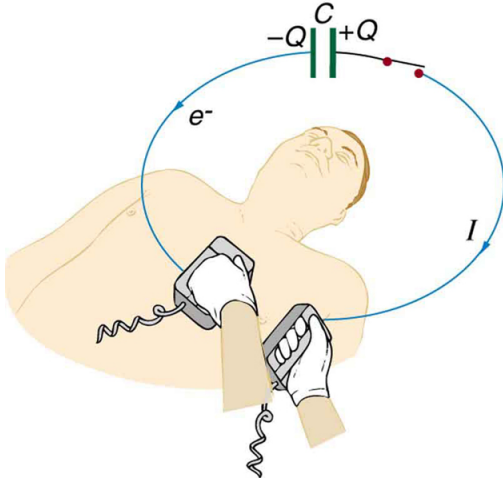
1.50ms

Exercise:**Problem:**

The 200-A current through a spark plug moves 0.300 mC of charge. How long does the spark last?

Exercise:**Problem:**

(a) A defibrillator sends a 6.00-A current through the chest of a patient by applying a 10,000-V potential as in the figure below. What is the resistance of the path? (b) The defibrillator paddles make contact with the patient through a conducting gel that greatly reduces the path resistance. Discuss the difficulties that would ensue if a larger voltage were used to produce the same current through the patient, but with the path having perhaps 50 times the resistance. (Hint: The current must be about the same, so a higher voltage would imply greater power. Use this equation for power: $P = I^2 R$.)



The capacitor in a defibrillation unit drives a current through the heart of a patient.

Solution:

(a) $1.67\text{k}\Omega$

(b) If a 50 times larger resistance existed, keeping the current about the same, the power would be increased by a factor of about 50 (based on the equation $P = I^2 R$), causing much more energy to be transferred to the skin, which could cause serious burns. The gel used reduces the resistance, and therefore reduces the power transferred to the skin.

Exercise:

Problem:

During open-heart surgery, a defibrillator can be used to bring a patient out of cardiac arrest. The resistance of the path is $500\ \Omega$ and a 10.0-mA current is needed. What voltage should be applied?

Exercise:

Problem:

(a) A defibrillator passes 12.0 A of current through the torso of a person for 0.0100 s. How much charge moves? (b) How many electrons pass through the wires connected to the patient? (See figure two problems earlier.)

Solution:

(a) 0.120 C

(b) 7.50×10^{17} electrons

Exercise:**Problem:**

A clock battery wears out after moving 10,000 C of charge through the clock at a rate of 0.500 mA. (a) How long did the clock run? (b) How many electrons per second flowed?

Exercise:**Problem:**

The batteries of a submerged non-nuclear submarine supply 1000 A at full speed ahead. How long does it take to move Avogadro's number (6.02×10^{23}) of electrons at this rate?

Solution:

96.3 s

Exercise:

Problem:

Electron guns are used in X-ray tubes. The electrons are accelerated through a relatively large voltage and directed onto a metal target, producing X-rays. (a) How many electrons per second strike the target if the current is 0.500 mA? (b) What charge strikes the target in 0.750 s?

Exercise:**Problem:**

A large cyclotron directs a beam of He^{++} nuclei onto a target with a beam current of 0.250 mA. (a) How many He^{++} nuclei per second is this? (b) How long does it take for 1.00 C to strike the target? (c) How long before 1.00 mol of He^{++} nuclei strike the target?

Solution:

(a) $7.81 \times 10^{14} \text{ He}^{++} \text{ nuclei/s}$

(b) $4.00 \times 10^3 \text{ s}$

(c) $7.71 \times 10^8 \text{ s}$

Exercise:**Problem:**

Repeat the above example on [\[link\]](#), but for a wire made of silver and given there is one free electron per silver atom.

Exercise:**Problem:**

Using the results of the above example on [\[link\]](#), find the drift velocity in a copper wire of twice the diameter and carrying 20.0 A.

Solution:

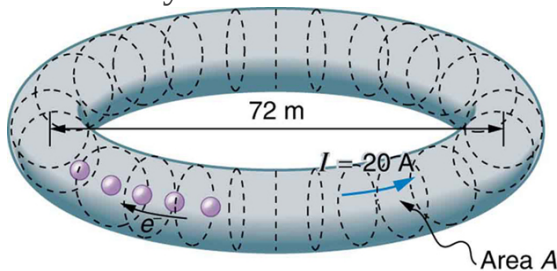
$$-1.13 \times 10^{-4} \text{ m/s}$$

Exercise:**Problem:**

A 14-gauge copper wire has a diameter of 1.628 mm. What magnitude current flows when the drift velocity is 1.00 mm/s? (See above example on [\[link\]](#) for useful information.)

Exercise:**Problem:**

SPEAR, a storage ring about 72.0 m in diameter at the Stanford Linear Accelerator (closed in 2009), has a 20.0-A circulating beam of electrons that are moving at nearly the speed of light. (See [\[link\]](#).) How many electrons are in the beam?



Electrons circulating in the storage ring called SPEAR constitute a 20.0-A current. Because they travel close to the speed of light, each electron completes many orbits in each second.

Solution:

$$9.42 \times 10^{13} \text{ electrons}$$

Glossary

electric current

the rate at which charge flows, $I = \Delta Q / \Delta t$

ampere

(amp) the SI unit for current; $1 \text{ A} = 1 \text{ C/s}$

drift velocity

the average velocity at which free charges flow in response to an electric field

Ohm's Law: Resistance and Simple Circuits

- Explain the origin of Ohm's law.
- Calculate voltages, currents, or resistances with Ohm's law.
- Explain what an ohmic material is.
- Describe a simple circuit.

What drives current? We can think of various devices—such as batteries, generators, wall outlets, and so on—which are necessary to maintain a current. All such devices create a potential difference and are loosely referred to as voltage sources. When a voltage source is connected to a conductor, it applies a potential difference V that creates an electric field. The electric field in turn exerts force on charges, causing current.

Ohm's Law

The current that flows through most substances is directly proportional to the voltage V applied to it. The German physicist Georg Simon Ohm (1787–1854) was the first to demonstrate experimentally that the current in a metal wire is *directly proportional to the voltage applied*:

Equation:

$$I \propto V.$$

This important relationship is known as **Ohm's law**. It can be viewed as a cause-and-effect relationship, with voltage the cause and current the effect. This is an empirical law like that for friction—an experimentally observed phenomenon. Such a linear relationship doesn't always occur.

Resistance and Simple Circuits

If voltage drives current, what impedes it? The electric property that impedes current (crudely similar to friction and air resistance) is called **resistance** R . Collisions of moving charges with atoms and molecules in a substance transfer energy to the substance and limit current. Resistance is defined as inversely proportional to current, or

Equation:

$$I \propto \frac{1}{R}.$$

Thus, for example, current is cut in half if resistance doubles. Combining the relationships of current to voltage and current to resistance gives

Equation:

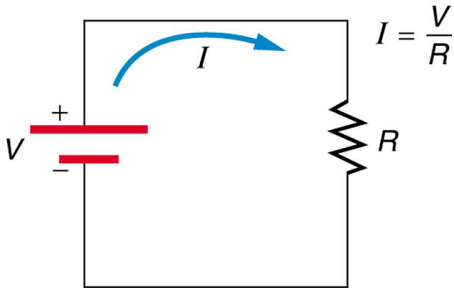
$$I = \frac{V}{R}.$$

This relationship is also called Ohm's law. Ohm's law in this form really defines resistance for certain materials. Ohm's law (like Hooke's law) is not universally valid. The many substances for which Ohm's law holds are called **ohmic**. These include good conductors like copper and aluminum, and some poor conductors under certain circumstances. Ohmic materials have a resistance R that is independent of voltage V and current I . An object that has simple resistance is called a *resistor*, even if its resistance is small. The unit for resistance is an **ohm** and is given the symbol Ω (upper case Greek omega). Rearranging $I = V/R$ gives $R = V/I$, and so the units of resistance are 1 ohm = 1 volt per ampere:

Equation:

$$1 \Omega = 1 \frac{V}{A}.$$

[\[link\]](#) shows the schematic for a simple circuit. A **simple circuit** has a single voltage source and a single resistor. The wires connecting the voltage source to the resistor can be assumed to have negligible resistance, or their resistance can be included in R .



A simple electric circuit in which a closed path for current to flow is supplied by conductors (usually metal wires) connecting a load to the terminals of a battery, represented by the red parallel lines.

The zigzag symbol represents the single resistor and includes any resistance in the connections to the voltage source.

Example:**Calculating Resistance: An Automobile Headlight**

What is the resistance of an automobile headlight through which 2.50 A flows when 12.0 V is applied to it?

Strategy

We can rearrange Ohm's law as stated by $I = V/R$ and use it to find the resistance.

Solution

Rearranging $I = V/R$ and substituting known values gives

Equation:

$$R = \frac{V}{I} = \frac{12.0 \text{ V}}{2.50 \text{ A}} = 4.80 \Omega.$$

Discussion

This is a relatively small resistance, but it is larger than the cold resistance of the headlight. As we shall see in [Resistance and Resistivity](#), resistance usually increases with temperature, and so the bulb has a lower resistance when it is first switched on and will draw considerably more current during its brief warm-up period.

Resistances range over many orders of magnitude. Some ceramic insulators, such as those used to support power lines, have resistances of $10^{12} \Omega$ or more. A dry person may have a hand-to-foot resistance of $10^5 \Omega$, whereas the resistance of the human heart is about $10^3 \Omega$. A meter-long piece of large-diameter copper wire may have a resistance of $10^{-5} \Omega$, and superconductors have no resistance at all (they are non-ohmic). Resistance is related to the shape of an object and the material of which it is composed, as will be seen in [Resistance and Resistivity](#).

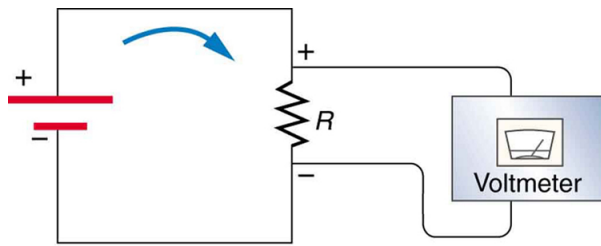
Additional insight is gained by solving $I = V/R$ for V , yielding

Equation:

$$V = IR.$$

This expression for V can be interpreted as the *voltage drop across a resistor produced by the flow of current I* . The phrase *IR drop* is often used for this voltage. For instance, the headlight in [\[link\]](#) has an IR drop of 12.0 V. If voltage is measured at various points in a circuit, it will be seen to increase at the voltage source and decrease at the resistor. Voltage is similar to fluid pressure. The voltage source is like a pump, creating a pressure difference, causing current—the flow of charge. The resistor is like a pipe that reduces pressure and limits flow because of its resistance. Conservation of energy has important consequences here. The voltage source supplies

energy (causing an electric field and a current), and the resistor converts it to another form (such as thermal energy). In a simple circuit (one with a single simple resistor), the voltage supplied by the source equals the voltage drop across the resistor, since $PE = q\Delta V$, and the same q flows through each. Thus the energy supplied by the voltage source and the energy converted by the resistor are equal. (See [\[link\]](#).)



$$V = IR = 18 \text{ V}$$

The voltage drop across a resistor in a simple circuit equals the voltage output of the battery.

Note:**Making Connections: Conservation of Energy**

In a simple electrical circuit, the sole resistor converts energy supplied by the source into another form. Conservation of energy is evidenced here by the fact that all of the energy supplied by the source is converted to another form by the resistor alone. We will find that conservation of energy has other important applications in circuits and is a powerful tool in circuit analysis.

Note:

PhET Explorations: Ohm's Law

See how the equation form of Ohm's law relates to a simple circuit. Adjust the voltage and resistance, and see the current change according to Ohm's law. The sizes of the symbols in the equation change to match the circuit diagram.

https://phet.colorado.edu/sims/html/ohms-law/latest/ohms-law_en.html

Section Summary

- A simple circuit *is* one in which there is a single voltage source and a single resistance.
- One statement of Ohm's law gives the relationship between current I , voltage V , and resistance R in a simple circuit to be $I = \frac{V}{R}$.
- Resistance has units of ohms (Ω), related to volts and amperes by $1 \Omega = 1 \text{ V/A}$.
- There is a voltage or IR drop across a resistor, caused by the current flowing through it, given by $V = IR$.

Conceptual Questions

Exercise:

Problem:

The IR drop across a resistor means that there is a change in potential or voltage across the resistor. Is there any change in current as it passes through a resistor? Explain.

Exercise:

Problem:

How is the IR drop in a resistor similar to the pressure drop in a fluid flowing through a pipe?

Problems & Exercises

Exercise:**Problem:**

What current flows through the bulb of a 3.00-V flashlight when its hot resistance is $3.60\ \Omega$?

Solution:

0.833 A

Exercise:**Problem:**

Calculate the effective resistance of a pocket calculator that has a 1.35-V battery and through which 0.200 mA flows.

Exercise:**Problem:**

What is the effective resistance of a car's starter motor when 150 A flows through it as the car battery applies 11.0 V to the motor?

Solution:

$7.33 \times 10^{-2}\ \Omega$

Exercise:**Problem:**

How many volts are supplied to operate an indicator light on a DVD player that has a resistance of $140\ \Omega$, given that 25.0 mA passes through it?

Exercise:

Problem:

(a) Find the voltage drop in an extension cord having a $0.0600\text{-}\Omega$ resistance and through which 5.00 A is flowing. (b) A cheaper cord utilizes thinner wire and has a resistance of $0.300\text{ }\Omega$. What is the voltage drop in it when 5.00 A flows? (c) Why is the voltage to whatever appliance is being used reduced by this amount? What is the effect on the appliance?

Solution:

(a) 0.300 V

(b) 1.50 V

(c) The voltage supplied to whatever appliance is being used is reduced because the total voltage drop from the wall to the final output of the appliance is fixed. Thus, if the voltage drop across the extension cord is large, the voltage drop across the appliance is significantly decreased, so the power output by the appliance can be significantly decreased, reducing the ability of the appliance to work properly.

Exercise:**Problem:**

A power transmission line is hung from metal towers with glass insulators having a resistance of $1.00 \times 10^9\text{ }\Omega$. What current flows through the insulator if the voltage is 200 kV ? (Some high-voltage lines are DC.)

Glossary**Ohm's law**

an empirical relation stating that the current I is proportional to the potential difference V , $\propto V$; it is often written as $I = V/R$, where R is the resistance

resistance

the electric property that impedes current; for ohmic materials, it is the ratio of voltage to current, $R = V/I$

ohm

the unit of resistance, given by $1\Omega = 1 \text{ V/A}$

ohmic

a type of a material for which Ohm's law is valid

simple circuit

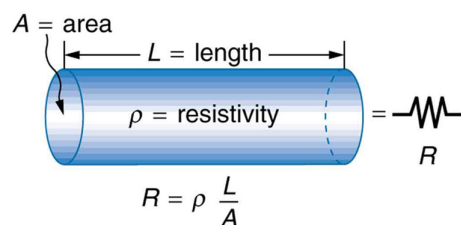
a circuit with a single voltage source and a single resistor

Resistance and Resistivity

- Explain the concept of resistivity.
- Use resistivity to calculate the resistance of specified configurations of material.
- Use the thermal coefficient of resistivity to calculate the change of resistance with temperature.

Material and Shape Dependence of Resistance

The resistance of an object depends on its shape and the material of which it is composed. The cylindrical resistor in [\[link\]](#) is easy to analyze, and, by so doing, we can gain insight into the resistance of more complicated shapes. As you might expect, the cylinder's electric resistance R is directly proportional to its length L , similar to the resistance of a pipe to fluid flow. The longer the cylinder, the more collisions charges will make with its atoms. The greater the diameter of the cylinder, the more current it can carry (again similar to the flow of fluid through a pipe). In fact, R is inversely proportional to the cylinder's cross-sectional area A .



A uniform cylinder of length L and cross-sectional area A . Its resistance to the flow of current is similar to the resistance posed by a pipe to fluid flow.

The longer the cylinder, the greater its

resistance. The larger its cross-sectional area A , the smaller its resistance.

For a given shape, the resistance depends on the material of which the object is composed. Different materials offer different resistance to the flow of charge. We define the **resistivity** ρ of a substance so that the **resistance** R of an object is directly proportional to ρ . Resistivity ρ is an *intrinsic* property of a material, independent of its shape or size. The resistance R of a uniform cylinder of length L , of cross-sectional area A , and made of a material with resistivity ρ , is

Equation:

$$R = \frac{\rho L}{A}.$$

[\[link\]](#) gives representative values of ρ . The materials listed in the table are separated into categories of conductors, semiconductors, and insulators, based on broad groupings of resistivities. Conductors have the smallest resistivities, and insulators have the largest; semiconductors have intermediate resistivities. Conductors have varying but large free charge densities, whereas most charges in insulators are bound to atoms and are not free to move. Semiconductors are intermediate, having far fewer free charges than conductors, but having properties that make the number of free charges depend strongly on the type and amount of impurities in the semiconductor. These unique properties of semiconductors are put to use in modern electronics, as will be explored in later chapters.

Material	Resistivity ρ ($\Omega \cdot \text{m}$)
<i>Conductors</i>	
Silver	1.59×10^{-8}
Copper	1.72×10^{-8}
Gold	2.44×10^{-8}
Aluminum	2.65×10^{-8}
Tungsten	5.6×10^{-8}
Iron	9.71×10^{-8}
Platinum	10.6×10^{-8}
Steel	20×10^{-8}
Lead	22×10^{-8}

Material	Resistivity ρ ($\Omega \cdot \text{m}$)
Manganin (Cu, Mn, Ni alloy)	44×10^{-8}
Constantan (Cu, Ni alloy)	49×10^{-8}
Mercury	96×10^{-8}
Nichrome (Ni, Fe, Cr alloy)	100×10^{-8}
<i>Semiconductors</i> [footnote] Values depend strongly on amounts and types of impurities	
Carbon (pure)	3.5×10^{-5}
Carbon	$(3.5 - 60) \times 10^{-5}$
Germanium (pure)	600×10^{-3}
Germanium	$(1 - 600) \times 10^{-3}$

Material	Resistivity ρ ($\Omega \cdot \text{m}$)
Silicon (pure)	2300
Silicon	0.1–2300
<i>Insulators</i>	
Amber	5×10^{14}
Glass	$10^9 - 10^{14}$
Lucite	$>10^{13}$
Mica	$10^{11} - 10^{15}$
Quartz (fused)	75×10^{16}
Rubber (hard)	$10^{13} - 10^{16}$
Sulfur	10^{15}

Material	Resistivity ρ ($\Omega \cdot \text{m}$)
Teflon	$>10^{13}$
Wood	$10^8 - 10^{11}$

Resistivities ρ of Various materials at 20°C

Example:

Calculating Resistor Diameter: A Headlight Filament

A car headlight filament is made of tungsten and has a cold resistance of 0.350Ω . If the filament is a cylinder 4.00 cm long (it may be coiled to save space), what is its diameter?

Strategy

We can rearrange the equation $R = \frac{\rho L}{A}$ to find the cross-sectional area A of the filament from the given information. Then its diameter can be found by assuming it has a circular cross-section.

Solution

The cross-sectional area, found by rearranging the expression for the resistance of a cylinder given in $R = \frac{\rho L}{A}$, is

Equation:

$$A = \frac{\rho L}{R}.$$

Substituting the given values, and taking ρ from [\[link\]](#), yields

Equation:

$$\begin{aligned} A &= \frac{(5.6 \times 10^{-8} \Omega \cdot \text{m})(4.00 \times 10^{-2} \text{ m})}{0.350 \Omega} \\ &= 6.40 \times 10^{-9} \text{ m}^2. \end{aligned}$$

The area of a circle is related to its diameter D by

Equation:

$$A = \frac{\pi D^2}{4}.$$

Solving for the diameter D , and substituting the value found for A , gives

Equation:

$$\begin{aligned} D &= 2\left(\frac{A}{\pi}\right)^{\frac{1}{2}} = 2\left(\frac{6.40 \times 10^{-9} \text{ m}^2}{3.14}\right)^{\frac{1}{2}} \\ &= 9.0 \times 10^{-5} \text{ m}. \end{aligned}$$

Discussion

The diameter is just under a tenth of a millimeter. It is quoted to only two digits, because ρ is known to only two digits.

Temperature Variation of Resistance

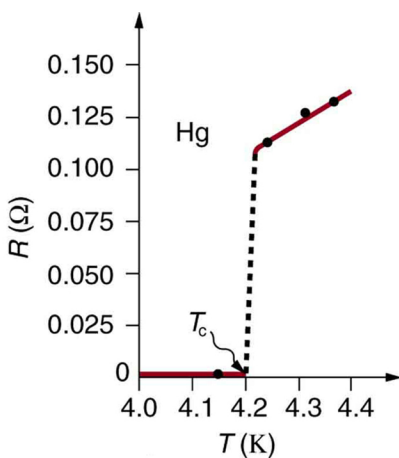
The resistivity of all materials depends on temperature. Some even become superconductors (zero resistivity) at very low temperatures. (See [\[link\]](#).) Conversely, the resistivity of conductors increases with increasing temperature. Since the atoms vibrate more rapidly and over larger distances at higher temperatures, the electrons moving through a metal make more collisions, effectively making the resistivity higher. Over relatively small temperature changes (about 100°C or less), resistivity ρ varies with temperature change ΔT as expressed in the following equation

Equation:

$$\rho = \rho_0(1 + \alpha\Delta T),$$

where ρ_0 is the original resistivity and α is the **temperature coefficient of resistivity**. (See the values of α in [\[link\]](#) below.) For larger temperature changes, α may vary or a nonlinear equation may be needed to find ρ . Note

that α is positive for metals, meaning their resistivity increases with temperature. Some alloys have been developed specifically to have a small temperature dependence. Manganin (which is made of copper, manganese and nickel), for example, has α close to zero (to three digits on the scale in [\[link\]](#)), and so its resistivity varies only slightly with temperature. This is useful for making a temperature-independent resistance standard, for example.



The resistance of a sample of mercury is zero at very low temperatures—it is a superconductor up to about 4.2 K. Above that critical temperature, its resistance makes a sudden jump and then increases nearly linearly with temperature.

Material	Coefficient $\alpha(1/^{\circ}\text{C})$ [footnote] Values at 20°C.
<i>Conductors</i>	
Silver	3.8×10^{-3}
Copper	3.9×10^{-3}
Gold	3.4×10^{-3}
Aluminum	3.9×10^{-3}
Tungsten	4.5×10^{-3}
Iron	5.0×10^{-3}
Platinum	3.93×10^{-3}
Lead	3.9×10^{-3}
Manganin (Cu, Mn, Ni alloy)	0.000×10^{-3}

Material	Coefficient $\alpha(1/^{\circ}\text{C})$ [footnote] Values at 20°C.
Constantan (Cu, Ni alloy)	0.002×10^{-3}
Mercury	0.89×10^{-3}
Nichrome (Ni, Fe, Cr alloy)	0.4×10^{-3}
<i>Semiconductors</i>	
Carbon (pure)	-0.5×10^{-3}
Germanium (pure)	-50×10^{-3}
Silicon (pure)	-70×10^{-3}

Temperature Coefficients of Resistivity α

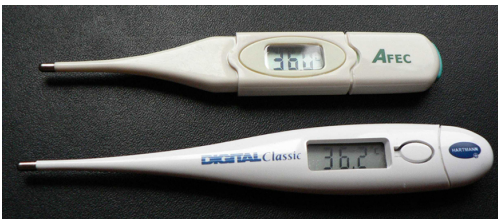
Note also that α is negative for the semiconductors listed in [\[link\]](#), meaning that their resistivity decreases with increasing temperature. They become better conductors at higher temperature, because increased thermal agitation increases the number of free charges available to carry current. This property of decreasing ρ with temperature is also related to the type and amount of impurities present in the semiconductors.

The resistance of an object also depends on temperature, since R_0 is directly proportional to ρ . For a cylinder we know $R = \rho L/A$, and so, if L and A do not change greatly with temperature, R will have the same temperature dependence as ρ . (Examination of the coefficients of linear expansion shows them to be about two orders of magnitude less than typical temperature coefficients of resistivity, and so the effect of temperature on L and A is about two orders of magnitude less than on ρ .) Thus,

Equation:

$$R = R_0(1 + \alpha\Delta T)$$

is the temperature dependence of the resistance of an object, where R_0 is the original resistance and R is the resistance after a temperature change ΔT . Numerous thermometers are based on the effect of temperature on resistance. (See [\[link\]](#).) One of the most common is the thermistor, a semiconductor crystal with a strong temperature dependence, the resistance of which is measured to obtain its temperature. The device is small, so that it quickly comes into thermal equilibrium with the part of a person it touches.



These familiar
thermometers are based
on the automated
measurement of a
thermistor's temperature-
dependent resistance.
(credit: Biol, Wikimedia
Commons)

Example:**Calculating Resistance: Hot-Filament Resistance**

Although caution must be used in applying $\rho = \rho_0(1 + \alpha\Delta T)$ and $R = R_0(1 + \alpha\Delta T)$ for temperature changes greater than 100°C, for tungsten the equations work reasonably well for very large temperature changes. What, then, is the resistance of the tungsten filament in the previous example if its temperature is increased from room temperature (20°C) to a typical operating temperature of 2850°C?

Strategy

This is a straightforward application of $R = R_0(1 + \alpha\Delta T)$, since the original resistance of the filament was given to be $R_0 = 0.350 \, \Omega$, and the temperature change is $\Delta T = 2830^\circ\text{C}$.

Solution

The hot resistance R is obtained by entering known values into the above equation:

Equation:

$$\begin{aligned} R &= R_0(1 + \alpha\Delta T) \\ &= (0.350 \, \Omega)[1 + (4.5 \times 10^{-3}/^\circ\text{C})(2830^\circ\text{C})] \\ &= 4.8 \, \Omega. \end{aligned}$$

Discussion

This value is consistent with the headlight resistance example in [Ohm's Law: Resistance and Simple Circuits](#).

Note:**PhET Explorations: Resistance in a Wire**

Learn about the physics of resistance in a wire. Change its resistivity, length, and area to see how they affect the wire's resistance. The sizes of the symbols in the equation change along with the diagram of a wire.

Section Summary

- The resistance R of a cylinder of length L and cross-sectional area A is $R = \frac{\rho L}{A}$, where ρ is the resistivity of the material.
- Values of ρ in [\[link\]](#) show that materials fall into three groups—*conductors, semiconductors, and insulators*.
- Temperature affects resistivity; for relatively small temperature changes ΔT , resistivity is $\rho = \rho_0(1 + \alpha\Delta T)$, where ρ_0 is the original resistivity and α is the temperature coefficient of resistivity.
- [\[link\]](#) gives values for α , the temperature coefficient of resistivity.
- The resistance R of an object also varies with temperature: $R = R_0(1 + \alpha\Delta T)$, where R_0 is the original resistance, and R is the resistance after the temperature change.

Conceptual Questions

Exercise:

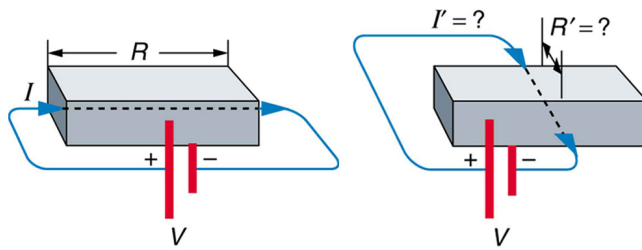
Problem:

In which of the three semiconducting materials listed in [\[link\]](#) do impurities supply free charges? (Hint: Examine the range of resistivity for each and determine whether the pure semiconductor has the higher or lower conductivity.)

Exercise:

Problem:

Does the resistance of an object depend on the path current takes through it? Consider, for example, a rectangular bar—is its resistance the same along its length as across its width? (See [\[link\]](#).)



Does current taking two different paths through the same object encounter different resistance?

Exercise:

Problem:

If aluminum and copper wires of the same length have the same resistance, which has the larger diameter? Why?

Exercise:

Problem:

Explain why $R = R_0(1 + \alpha\Delta T)$ for the temperature variation of the resistance R of an object is not as accurate as $\rho = \rho_0(1 + \alpha\Delta T)$, which gives the temperature variation of resistivity ρ .

Problems & Exercises

Exercise:

Problem:

What is the resistance of a 20.0-m-long piece of 12-gauge copper wire having a 2.053-mm diameter?

Solution:

0.104 Ω

Exercise:**Problem:**

The diameter of 0-gauge copper wire is 8.252 mm. Find the resistance of a 1.00-km length of such wire used for power transmission.

Exercise:**Problem:**

If the 0.100-mm diameter tungsten filament in a light bulb is to have a resistance of $0.200\ \Omega$ at 20.0°C , how long should it be?

Solution:

$$2.8 \times 10^{-2}\ \text{m}$$

Exercise:**Problem:**

Find the ratio of the diameter of aluminum to copper wire, if they have the same resistance per unit length (as they might in household wiring).

Exercise:**Problem:**

What current flows through a 2.54-cm-diameter rod of pure silicon that is 20.0 cm long, when $1.00 \times 10^3\ \text{V}$ is applied to it? (Such a rod may be used to make nuclear-particle detectors, for example.)

Solution:

$$1.10 \times 10^{-3}\ \text{A}$$

Exercise:

Problem:

(a) To what temperature must you raise a copper wire, originally at 20.0°C , to double its resistance, neglecting any changes in dimensions? (b) Does this happen in household wiring under ordinary circumstances?

Exercise:**Problem:**

A resistor made of Nichrome wire is used in an application where its resistance cannot change more than 1.00% from its value at 20.0°C . Over what temperature range can it be used?

Solution:

-5°C to 45°C

Exercise:**Problem:**

Of what material is a resistor made if its resistance is 40.0% greater at 100°C than at 20.0°C ?

Exercise:**Problem:**

An electronic device designed to operate at any temperature in the range from -10.0°C to 55.0°C contains pure carbon resistors. By what factor does their resistance increase over this range?

Solution:

1.03

Exercise:

Problem:

(a) Of what material is a wire made, if it is 25.0 m long with a 0.100 mm diameter and has a resistance of $77.7\ \Omega$ at 20.0°C ? (b) What is its resistance at 150°C ?

Exercise:**Problem:**

Assuming a constant temperature coefficient of resistivity, what is the maximum percent decrease in the resistance of a constantan wire starting at 20.0°C ?

Solution:

0.06%

Exercise:**Problem:**

A wire is drawn through a die, stretching it to four times its original length. By what factor does its resistance increase?

Exercise:**Problem:**

A copper wire has a resistance of $0.500\ \Omega$ at 20.0°C , and an iron wire has a resistance of $0.525\ \Omega$ at the same temperature. At what temperature are their resistances equal?

Solution:

-17°C

Exercise:

Problem:

(a) Digital medical thermometers determine temperature by measuring the resistance of a semiconductor device called a thermistor (which has $\alpha = -0.0600/^{\circ}\text{C}$) when it is at the same temperature as the patient. What is a patient's temperature if the thermistor's resistance at that temperature is 82.0% of its value at 37.0°C (normal body temperature)? (b) The negative value for α may not be maintained for very low temperatures. Discuss why and whether this is the case here. (Hint: Resistance can't become negative.)

Exercise:**Problem: Integrated Concepts**

(a) Redo [\[link\]](#) taking into account the thermal expansion of the tungsten filament. You may assume a thermal expansion coefficient of $12 \times 10^{-6}/^{\circ}\text{C}$. (b) By what percentage does your answer differ from that in the example?

Solution:

(a) $4.7\ \Omega$ (total)

(b) 3.0% decrease

Exercise:**Problem: Unreasonable Results**

(a) To what temperature must you raise a resistor made of constantan to double its resistance, assuming a constant temperature coefficient of resistivity? (b) To cut it in half? (c) What is unreasonable about these results? (d) Which assumptions are unreasonable, or which premises are inconsistent?

Glossary

resistivity

an intrinsic property of a material, independent of its shape or size,
directly proportional to the resistance, denoted by ρ

temperature coefficient of resistivity

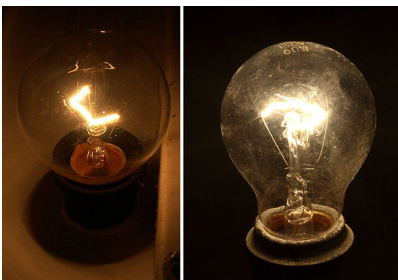
an empirical quantity, denoted by α , which describes the change in
resistance or resistivity of a material with temperature

Electric Power and Energy

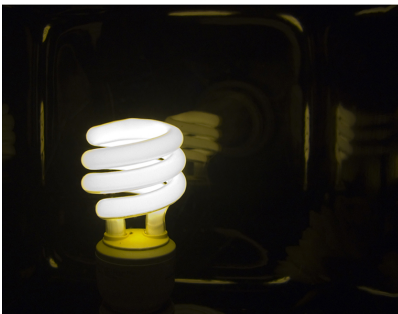
- Calculate the power dissipated by a resistor and power supplied by a power supply.
- Calculate the cost of electricity under various circumstances.

Power in Electric Circuits

Power is associated by many people with electricity. Knowing that power is the rate of energy use or energy conversion, what is the expression for **electric power**? Power transmission lines might come to mind. We also think of lightbulbs in terms of their power ratings in watts. Let us compare a 25-W bulb with a 60-W bulb. (See [\[link\]](#)(a).) Since both operate on the same voltage, the 60-W bulb must draw more current to have a greater power rating. Thus the 60-W bulb's resistance must be lower than that of a 25-W bulb. If we increase voltage, we also increase power. For example, when a 25-W bulb that is designed to operate on 120 V is connected to 240 V, it briefly glows very brightly and then burns out. Precisely how are voltage, current, and resistance related to electric power?



(a)



(b)

(a) Which of these lightbulbs, the 25-W bulb (upper left) or the 60-W bulb (upper right), has the higher resistance? Which draws more current? Which uses the most energy? Can you tell from the color that the 25-W filament is cooler? Is the brighter bulb a different color and if so why?

(credits: Dickbauch, Wikimedia Commons; Greg Westfall, Flickr) (b)

This compact fluorescent light (CFL) puts out the same intensity of light as the 60-W bulb, but at 1/4 to 1/10 the input power. (credit: dbgg1979, Flickr)

Electric energy depends on both the voltage involved and the charge moved. This is expressed most simply as $PE = qV$, where q is the charge moved and V is the voltage (or more precisely, the potential difference the

charge moves through). Power is the rate at which energy is moved, and so electric power is

Equation:

$$P = \frac{PE}{t} = \frac{qV}{t}.$$

Recognizing that current is $I = q/t$ (note that $\Delta t = t$ here), the expression for power becomes

Equation:

$$P = IV.$$

Electric power (P) is simply the product of current times voltage. Power has familiar units of watts. Since the SI unit for potential energy (PE) is the joule, power has units of joules per second, or watts. Thus, $1 \text{ A} \cdot \text{V} = 1 \text{ W}$. For example, cars often have one or more auxiliary power outlets with which you can charge a cell phone or other electronic devices. These outlets may be rated at 20 A, so that the circuit can deliver a maximum power $P = IV = (20 \text{ A})(12 \text{ V}) = 240 \text{ W}$. In some applications, electric power may be expressed as volt-amperes or even kilovolt-amperes ($1 \text{ kA} \cdot \text{V} = 1 \text{ kW}$).

To see the relationship of power to resistance, we combine Ohm's law with $P = IV$. Substituting $I = V/R$ gives $P = (V/R)V = V^2/R$. Similarly, substituting $V = IR$ gives $P = I(IR) = I^2R$. Three expressions for electric power are listed together here for convenience:

Equation:

$$P = IV$$

Equation:

$$P = \frac{V^2}{R}$$

Equation:

$$P = I^2 R.$$

Note that the first equation is always valid, whereas the other two can be used only for resistors. In a simple circuit, with one voltage source and a single resistor, the power supplied by the voltage source and that dissipated by the resistor are identical. (In more complicated circuits, P can be the power dissipated by a single device and not the total power in the circuit.)

Different insights can be gained from the three different expressions for electric power. For example, $P = V^2/R$ implies that the lower the resistance connected to a given voltage source, the greater the power delivered. Furthermore, since voltage is squared in $P = V^2/R$, the effect of applying a higher voltage is perhaps greater than expected. Thus, when the voltage is doubled to a 25-W bulb, its power nearly quadruples to about 100 W, burning it out. If the bulb's resistance remained constant, its power would be exactly 100 W, but at the higher temperature its resistance is higher, too.

Example:**Calculating Power Dissipation and Current: Hot and Cold Power**

(a) Consider the examples given in [Ohm's Law: Resistance and Simple Circuits](#) and [Resistance and Resistivity](#). Then find the power dissipated by the car headlight in these examples, both when it is hot and when it is cold.
(b) What current does it draw when cold?

Strategy for (a)

For the hot headlight, we know voltage and current, so we can use $P = IV$ to find the power. For the cold headlight, we know the voltage and resistance, so we can use $P = V^2/R$ to find the power.

Solution for (a)

Entering the known values of current and voltage for the hot headlight, we obtain

Equation:

$$P = IV = (2.50 \text{ A})(12.0 \text{ V}) = 30.0 \text{ W}.$$

The cold resistance was $0.350\ \Omega$, and so the power it uses when first switched on is

Equation:

$$P = \frac{V^2}{R} = \frac{(12.0\ \text{V})^2}{0.350\ \Omega} = 411\ \text{W}.$$

Discussion for (a)

The 30 W dissipated by the hot headlight is typical. But the 411 W when cold is surprisingly higher. The initial power quickly decreases as the bulb's temperature increases and its resistance increases.

Strategy and Solution for (b)

The current when the bulb is cold can be found several different ways. We rearrange one of the power equations, $P = I^2 R$, and enter known values, obtaining

Equation:

$$I = \sqrt{\frac{P}{R}} = \sqrt{\frac{411\ \text{W}}{0.350\ \Omega}} = 34.3\ \text{A}.$$

Discussion for (b)

The cold current is remarkably higher than the steady-state value of 2.50 A, but the current will quickly decline to that value as the bulb's temperature increases. Most fuses and circuit breakers (used to limit the current in a circuit) are designed to tolerate very high currents briefly as a device comes on. In some cases, such as with electric motors, the current remains high for several seconds, necessitating special "slow blow" fuses.

The Cost of Electricity

The more electric appliances you use and the longer they are left on, the higher your electric bill. This familiar fact is based on the relationship between energy and power. You pay for the energy used. Since $P = E/t$, we see that

Equation:

$$E = Pt$$

is the energy used by a device using power P for a time interval t . For example, the more lightbulbs burning, the greater P used; the longer they are on, the greater t is. The energy unit on electric bills is the kilowatt-hour ($\text{kW} \cdot \text{h}$), consistent with the relationship $E = Pt$. It is easy to estimate the cost of operating electric appliances if you have some idea of their power consumption rate in watts or kilowatts, the time they are on in hours, and the cost per kilowatt-hour for your electric utility. Kilowatt-hours, like all other specialized energy units such as food calories, can be converted to joules. You can prove to yourself that $1 \text{ kW} \cdot \text{h} = 3.6 \times 10^6 \text{ J}$.

The electrical energy (E) used can be reduced either by reducing the time of use or by reducing the power consumption of that appliance or fixture. This will not only reduce the cost, but it will also result in a reduced impact on the environment. Improvements to lighting are some of the fastest ways to reduce the electrical energy used in a home or business. About 20% of a home's use of energy goes to lighting, while the number for commercial establishments is closer to 40%. Fluorescent lights are about four times more efficient than incandescent lights—this is true for both the long tubes and the compact fluorescent lights (CFL). (See [\[link\]](#)(b).) Thus, a 60-W incandescent bulb can be replaced by a 15-W CFL, which has the same brightness and color. CFLs have a bent tube inside a globe or a spiral-shaped tube, all connected to a standard screw-in base that fits standard incandescent light sockets. (Original problems with color, flicker, shape, and high initial investment for CFLs have been addressed in recent years.) The heat transfer from these CFLs is less, and they last up to 10 times longer. The significance of an investment in such bulbs is addressed in the next example. New white LED lights (which are clusters of small LED bulbs) are even more efficient (twice that of CFLs) and last 5 times longer than CFLs. However, their cost is still high.

Note:

Making Connections: Energy, Power, and Time

The relationship $E = Pt$ is one that you will find useful in many different contexts. The energy your body uses in exercise is related to the power level and duration of your activity, for example. The amount of heating by a power source is related to the power level and time it is applied. Even the radiation dose of an X-ray image is related to the power and time of exposure.

Example:

Calculating the Cost Effectiveness of Compact Fluorescent Lights (CFL)

If the cost of electricity in your area is 12 cents per kWh, what is the total cost (capital plus operation) of using a 60-W incandescent bulb for 1000 hours (the lifetime of that bulb) if the bulb cost 25 cents? (b) If we replace this bulb with a compact fluorescent light that provides the same light output, but at one-quarter the wattage, and which costs \$1.50 but lasts 10 times longer (10,000 hours), what will that total cost be?

Strategy

To find the operating cost, we first find the energy used in kilowatt-hours and then multiply by the cost per kilowatt-hour.

Solution for (a)

The energy used in kilowatt-hours is found by entering the power and time into the expression for energy:

Equation:

$$E = Pt = (60 \text{ W})(1000 \text{ h}) = 60,000 \text{ W} \cdot \text{h}.$$

In kilowatt-hours, this is

Equation:

$$E = 60.0 \text{ kW} \cdot \text{h}.$$

Now the electricity cost is

Equation:

$$\text{cost} = (60.0 \text{ kW} \cdot \text{h})(\$0.12/\text{kW} \cdot \text{h}) = \$7.20.$$

The total cost will be \$7.20 for 1000 hours (about one-half year at 5 hours per day).

Solution for (b)

Since the CFL uses only 15 W and not 60 W, the electricity cost will be $\$7.20/4 = \1.80 . The CFL will last 10 times longer than the incandescent, so that the investment cost will be 1/10 of the bulb cost for that time period of use, or $0.1(\$1.50) = \0.15 . Therefore, the total cost will be \$1.95 for 1000 hours.

Discussion

Therefore, it is much cheaper to use the CFLs, even though the initial investment is higher. The increased cost of labor that a business must include for replacing the incandescent bulbs more often has not been figured in here.

Note:

Making Connections: Take-Home Experiment—Electrical Energy Use Inventory

1) Make a list of the power ratings on a range of appliances in your home or room. Explain why something like a toaster has a higher rating than a digital clock. Estimate the energy consumed by these appliances in an average day (by estimating their time of use). Some appliances might only state the operating current. If the household voltage is 120 V, then use $P = IV$. 2) Check out the total wattage used in the rest rooms of your school's floor or building. (You might need to assume the long fluorescent lights in use are rated at 32 W.) Suppose that the building was closed all weekend and that these lights were left on from 6 p.m. Friday until 8 a.m. Monday. What would this oversight cost? How about for an entire year of weekends?

Section Summary

- Electric power P is the rate (in watts) that energy is supplied by a source or dissipated by a device.

- Three expressions for electrical power are
Equation:

$$P = IV,$$

Equation:

$$P = \frac{V^2}{R},$$

and

Equation:

$$P = I^2 R.$$

- The energy used by a device with a power P over a time t is $E = Pt$.

Conceptual Questions

Exercise:

Problem:

Why do incandescent lightbulbs grow dim late in their lives, particularly just before their filaments break?

Exercise:

Problem:

The power dissipated in a resistor is given by $P = V^2/R$, which means power decreases if resistance increases. Yet this power is also given by $P = I^2 R$, which means power increases if resistance increases. Explain why there is no contradiction here.

Problem Exercises

Exercise:

Problem:

What is the power of a 1.00×10^2 MV lightning bolt having a current of 2.00×10^4 A?

Solution:

$$2.00 \times 10^{12} \text{ W}$$

Exercise:**Problem:**

What power is supplied to the starter motor of a large truck that draws 250 A of current from a 24.0-V battery hookup?

Exercise:**Problem:**

A charge of 4.00 C of charge passes through a pocket calculator's solar cells in 4.00 h. What is the power output, given the calculator's voltage output is 3.00 V? (See [\[link\]](#).)



The strip of solar cells just above the keys of this calculator convert

light to electricity
to supply its energy
needs. (credit:
Evan-Amos,
Wikimedia
Commons)

Exercise:

Problem:

How many watts does a flashlight that has 6.00×10^2 C pass through it in 0.500 h use if its voltage is 3.00 V?

Exercise:

Problem:

Find the power dissipated in each of these extension cords: (a) an extension cord having a $0.0600\ \Omega$ resistance and through which 5.00 A is flowing; (b) a cheaper cord utilizing thinner wire and with a resistance of $0.300\ \Omega$.

Solution:

(a) 1.50 W

(b) 7.50 W

Exercise:

Problem:

Verify that the units of a volt-ampere are watts, as implied by the equation $P = IV$.

Exercise:

Problem:

Show that the units $1 \text{ V}^2/\Omega = 1 \text{ W}$, as implied by the equation $P = V^2/R$.

Solution:

$$\frac{V^2}{\Omega} = \frac{V^2}{V/A} = AV = \left(\frac{C}{s}\right)\left(\frac{J}{C}\right) = \frac{J}{s} = 1 \text{ W}$$

Exercise:**Problem:**

Show that the units $1 \text{ A}^2 \cdot \Omega = 1 \text{ W}$, as implied by the equation $P = I^2 R$.

Exercise:**Problem:**

Verify the energy unit equivalence that $1 \text{ kW} \cdot \text{h} = 3.60 \times 10^6 \text{ J}$.

Solution:

$$1 \text{ kW} \cdot \text{h} = \left(\frac{1 \times 10^3 \text{ J}}{1 \text{ s}}\right)(1 \text{ h})\left(\frac{3600 \text{ s}}{1 \text{ h}}\right) = 3.60 \times 10^6 \text{ J}$$

Exercise:**Problem:**

Electrons in an X-ray tube are accelerated through $1.00 \times 10^2 \text{ kV}$ and directed toward a target to produce X-rays. Calculate the power of the electron beam in this tube if it has a current of 15.0 mA .

Exercise:

Problem:

An electric water heater consumes 5.00 kW for 2.00 h per day. What is the cost of running it for one year if electricity costs 12.0 cents/kW · h? See [\[link\]](#).



On-demand electric hot water heater. Heat is supplied to water only when needed.
(credit: aviddavid, Flickr)

Solution:

\$438/y

Exercise:**Problem:**

With a 1200-W toaster, how much electrical energy is needed to make a slice of toast (cooking time = 1 minute)? At 9.0 cents/kW · h, how much does this cost?

Exercise:

Problem:

What would be the maximum cost of a CFL such that the total cost (investment plus operating) would be the same for both CFL and incandescent 60-W bulbs? Assume the cost of the incandescent bulb is 25 cents and that electricity costs 10 cents/kWh. Calculate the cost for 1000 hours, as in the cost effectiveness of CFL example.

Solution:

\$6.25

Exercise:**Problem:**

Some makes of older cars have 6.00-V electrical systems. (a) What is the hot resistance of a 30.0-W headlight in such a car? (b) What current flows through it?

Exercise:**Problem:**

Alkaline batteries have the advantage of putting out constant voltage until very nearly the end of their life. How long will an alkaline battery rated at 1.00 A · h and 1.58 V keep a 1.00-W flashlight bulb burning?

Solution:

1.58 h

Exercise:**Problem:**

A cauterizer, used to stop bleeding in surgery, puts out 2.00 mA at 15.0 kV. (a) What is its power output? (b) What is the resistance of the path?

Exercise:

Problem:

The average television is said to be on 6 hours per day. Estimate the yearly cost of electricity to operate 100 million TVs, assuming their power consumption averages 150 W and the cost of electricity averages 12.0 cents/kW · h.

Solution:

\$3.94 billion/year

Exercise:**Problem:**

An old lightbulb draws only 50.0 W, rather than its original 60.0 W, due to evaporative thinning of its filament. By what factor is its diameter reduced, assuming uniform thinning along its length? Neglect any effects caused by temperature differences.

Exercise:**Problem:**

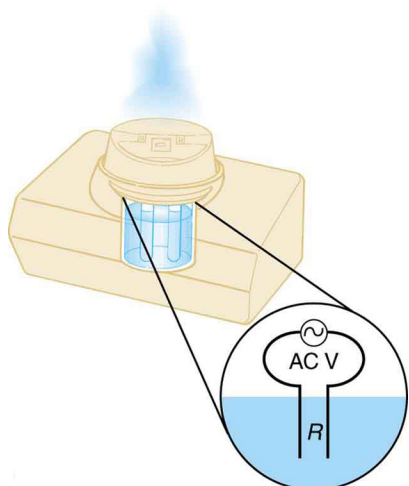
00-gauge copper wire has a diameter of 9.266 mm. Calculate the power loss in a kilometer of such wire when it carries 1.00×10^2 A.

Solution:

25.5 W

Exercise:**Problem: Integrated Concepts**

Cold vaporizers pass a current through water, evaporating it with only a small increase in temperature. One such home device is rated at 3.50 A and utilizes 120 V AC with 95.0% efficiency. (a) What is the vaporization rate in grams per minute? (b) How much water must you put into the vaporizer for 8.00 h of overnight operation? (See [\[link\]](#).)



This cold vaporizer
passes current
directly through
water, vaporizing it
directly with
relatively little
temperature
increase.

Exercise:

Problem: Integrated Concepts

(a) What energy is dissipated by a lightning bolt having a 20,000-A current, a voltage of 1.00×10^2 MV, and a length of 1.00 ms? (b) What mass of tree sap could be raised from 18.0°C to its boiling point and then evaporated by this energy, assuming sap has the same thermal characteristics as water?

Solution:

(a) 2.00×10^9 J

(b) 769 kg

Exercise:**Problem: Integrated Concepts**

What current must be produced by a 12.0-V battery-operated bottle warmer in order to heat 75.0 g of glass, 250 g of baby formula, and 3.00×10^2 g of aluminum from 20.0°C to 90.0°C in 5.00 min?

Exercise:**Problem: Integrated Concepts**

How much time is needed for a surgical cauterizer to raise the temperature of 1.00 g of tissue from 37.0°C to 100°C and then boil away 0.500 g of water, if it puts out 2.00 mA at 15.0 kV? Ignore heat transfer to the surroundings.

Solution:

45.0 s

Exercise:**Problem: Integrated Concepts**

Hydroelectric generators (see [\[link\]](#)) at Hoover Dam produce a maximum current of 8.00×10^3 A at 250 kV. (a) What is the power output? (b) The water that powers the generators enters and leaves the system at low speed (thus its kinetic energy does not change) but loses 160 m in altitude. How many cubic meters per second are needed, assuming 85.0% efficiency?



Hydroelectric generators
at the Hoover dam.
(credit: Jon Sullivan)

Exercise:

Problem: Integrated Concepts

(a) Assuming 95.0% efficiency for the conversion of electrical power by the motor, what current must the 12.0-V batteries of a 750-kg electric car be able to supply: (a) To accelerate from rest to 25.0 m/s in 1.00 min? (b) To climb a 2.00×10^2 -m-high hill in 2.00 min at a constant 25.0-m/s speed while exerting 5.00×10^2 N of force to overcome air resistance and friction? (c) To travel at a constant 25.0-m/s speed, exerting a 5.00×10^2 N force to overcome air resistance and friction? See [\[link\]](#).



This REVAi, an electric

car, gets recharged on a street in London. (credit: Frank Hebbert)

Solution:

(a) 343 A

(b) 2.17×10^3 A

(c) 1.10×10^3 A

Exercise:

Problem: Integrated Concepts

A light-rail commuter train draws 630 A of 650-V DC electricity when accelerating. (a) What is its power consumption rate in kilowatts? (b) How long does it take to reach 20.0 m/s starting from rest if its loaded mass is 5.30×10^4 kg, assuming 95.0% efficiency and constant power? (c) Find its average acceleration. (d) Discuss how the acceleration you found for the light-rail train compares to what might be typical for an automobile.

Exercise:

Problem: Integrated Concepts

(a) An aluminum power transmission line has a resistance of $0.0580 \Omega/\text{km}$. What is its mass per kilometer? (b) What is the mass per kilometer of a copper line having the same resistance? A lower resistance would shorten the heating time. Discuss the practical limits to speeding the heating by lowering the resistance.

Solution:

(a) $1.23 \times 10^3 \text{ kg}$

(b) $2.64 \times 10^3 \text{ kg}$

Exercise:

Problem: Integrated Concepts

(a) An immersion heater utilizing 120 V can raise the temperature of a 1.00×10^2 -g aluminum cup containing 350 g of water from 20.0°C to 95.0°C in 2.00 min. Find its resistance, assuming it is constant during the process. (b) A lower resistance would shorten the heating time. Discuss the practical limits to speeding the heating by lowering the resistance.

Exercise:

Problem: Integrated Concepts

(a) What is the cost of heating a hot tub containing 1500 kg of water from 10.0°C to 40.0°C, assuming 75.0% efficiency to account for heat transfer to the surroundings? The cost of electricity is 9 cents/kW · h. (b) What current was used by the 220-V AC electric heater, if this took 4.00 h?

Exercise:

Problem: Unreasonable Results

(a) What current is needed to transmit 1.00×10^2 MW of power at 480 V? (b) What power is dissipated by the transmission lines if they have a $1.00 - \Omega$ resistance? (c) What is unreasonable about this result? (d) Which assumptions are unreasonable, or which premises are inconsistent?

Solution:

(a) $2.08 \times 10^5 \text{ A}$

(b) 4.33×10^4 MW

(c) The transmission lines dissipate more power than they are supposed to transmit.

(d) A voltage of 480 V is unreasonably low for a transmission voltage. Long-distance transmission lines are kept at much higher voltages (often hundreds of kilovolts) to reduce power losses.

Exercise:

Problem: Unreasonable Results

(a) What current is needed to transmit 1.00×10^2 MW of power at 10.0 kV? (b) Find the resistance of 1.00 km of wire that would cause a 0.0100% power loss. (c) What is the diameter of a 1.00-km-long copper wire having this resistance? (d) What is unreasonable about these results? (e) Which assumptions are unreasonable, or which premises are inconsistent?

Exercise:

Problem: Construct Your Own Problem

Consider an electric immersion heater used to heat a cup of water to make tea. Construct a problem in which you calculate the needed resistance of the heater so that it increases the temperature of the water and cup in a reasonable amount of time. Also calculate the cost of the electrical energy used in your process. Among the things to be considered are the voltage used, the masses and heat capacities involved, heat losses, and the time over which the heating takes place. Your instructor may wish for you to consider a thermal safety switch (perhaps bimetallic) that will halt the process before damaging temperatures are reached in the immersion unit.

Glossary

electric power

the rate at which electrical energy is supplied by a source or dissipated by a device; it is the product of current times voltage

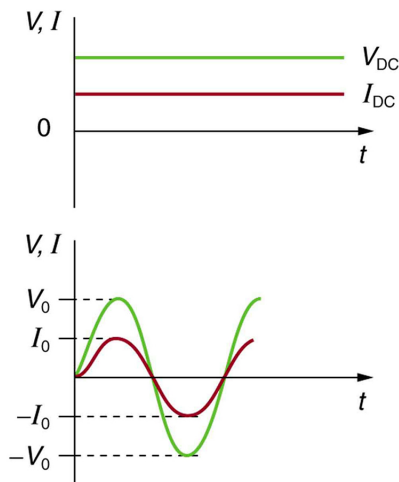
Alternating Current versus Direct Current

- Explain the differences and similarities between AC and DC current.
- Calculate rms voltage, current, and average power.
- Explain why AC current is used for power transmission.

Alternating Current

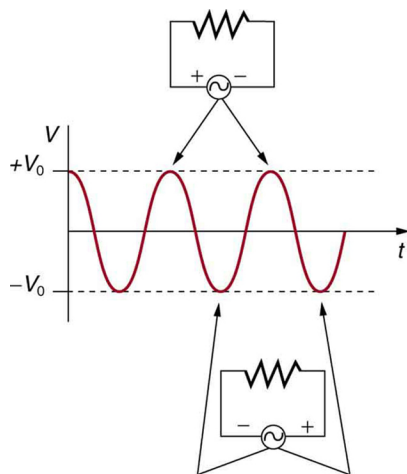
Most of the examples dealt with so far, and particularly those utilizing batteries, have constant voltage sources. Once the current is established, it is thus also a constant. **Direct current** (DC) is the flow of electric charge in only one direction. It is the steady state of a constant-voltage circuit. Most well-known applications, however, use a time-varying voltage source.

Alternating current (AC) is the flow of electric charge that periodically reverses direction. If the source varies periodically, particularly sinusoidally, the circuit is known as an alternating current circuit. Examples include the commercial and residential power that serves so many of our needs. [\[link\]](#) shows graphs of voltage and current versus time for typical DC and AC power. The AC voltages and frequencies commonly used in homes and businesses vary around the world.



(a) DC voltage and current are constant in time, once the

current is established. (b) A graph of voltage and current versus time for 60-Hz AC power. The voltage and current are sinusoidal and are in phase for a simple resistance circuit. The frequencies and peak voltages of AC sources differ greatly.



The potential difference V between the terminals of an AC voltage source fluctuates as

shown. The
mathematical
expression for V is
given by
 $V = V_0 \sin 2\pi ft$.

[\[link\]](#) shows a schematic of a simple circuit with an AC voltage source. The voltage between the terminals fluctuates as shown, with the **AC voltage** given by

Equation:

$$V = V_0 \sin 2\pi ft,$$

where V is the voltage at time t , V_0 is the peak voltage, and f is the frequency in hertz. For this simple resistance circuit, $I = V/R$, and so the **AC current** is

Equation:

$$I = I_0 \sin 2\pi ft,$$

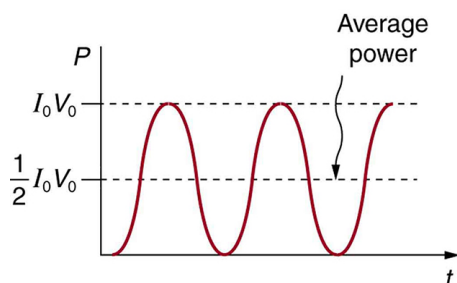
where I is the current at time t , and $I_0 = V_0/R$ is the peak current. For this example, the voltage and current are said to be in phase, as seen in [\[link\]](#)(b).

Current in the resistor alternates back and forth just like the driving voltage, since $I = V/R$. If the resistor is a fluorescent light bulb, for example, it brightens and dims 120 times per second as the current repeatedly goes through zero. A 120-Hz flicker is too rapid for your eyes to detect, but if you wave your hand back and forth between your face and a fluorescent light, you will see a stroboscopic effect evidencing AC. The fact that the light output fluctuates means that the power is fluctuating. The power supplied is $P = IV$. Using the expressions for I and V above, we see that the time dependence of power is $P = I_0 V_0 \sin^2 2\pi ft$, as shown in [\[link\]](#).

Note:**Making Connections: Take-Home Experiment—AC/DC Lights**

Wave your hand back and forth between your face and a fluorescent light bulb. Do you observe the same thing with the headlights on your car?

Explain what you observe. *Warning: Do not look directly at very bright light.*



AC power as a function of time. Since the voltage and current are in phase here, their product is non-negative and fluctuates between zero and $I_0 V_0$.

Average power is $(1/2)I_0 V_0$.

We are most often concerned with average power rather than its fluctuations—that 60-W light bulb in your desk lamp has an average power consumption of 60 W, for example. As illustrated in [\[link\]](#), the average power P_{ave} is

Equation:

$$P_{\text{ave}} = \frac{1}{2} I_0 V_0.$$

This is evident from the graph, since the areas above and below the $(1/2)I_0V_0$ line are equal, but it can also be proven using trigonometric identities. Similarly, we define an average or **rms current** I_{rms} and average or **rms voltage** V_{rms} to be, respectively,

Equation:

$$I_{\text{rms}} = \frac{I_0}{\sqrt{2}}$$

and

Equation:

$$V_{\text{rms}} = \frac{V_0}{\sqrt{2}}.$$

where rms stands for root mean square, a particular kind of average. In general, to obtain a root mean square, the particular quantity is squared, its mean (or average) is found, and the square root is taken. This is useful for AC, since the average value is zero. Now,

Equation:

$$P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}},$$

which gives

Equation:

$$P_{\text{ave}} = \frac{I_0}{\sqrt{2}} \cdot \frac{V_0}{\sqrt{2}} = \frac{1}{2} I_0 V_0,$$

as stated above. It is standard practice to quote I_{rms} , V_{rms} , and P_{ave} rather than the peak values. For example, most household electricity is 120 V AC, which means that V_{rms} is 120 V. The common 10-A circuit breaker will interrupt a sustained I_{rms} greater than 10 A. Your 1.0-kW microwave oven

consumes $P_{\text{ave}} = 1.0 \text{ kW}$, and so on. You can think of these rms and average values as the equivalent DC values for a simple resistive circuit.

To summarize, when dealing with AC, Ohm's law and the equations for power are completely analogous to those for DC, but rms and average values are used for AC. Thus, for AC, Ohm's law is written

Equation:

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{R}.$$

The various expressions for AC power P_{ave} are

Equation:

$$P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}},$$

Equation:

$$P_{\text{ave}} = \frac{V_{\text{rms}}^2}{R},$$

and

Equation:

$$P_{\text{ave}} = I_{\text{rms}}^2 R.$$

Example:

Peak Voltage and Power for AC

(a) What is the value of the peak voltage for 120-V AC power? (b) What is the peak power consumption rate of a 60.0-W AC light bulb?

Strategy

We are told that V_{rms} is 120 V and P_{ave} is 60.0 W. We can use $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$ to find the peak voltage, and we can manipulate the definition of power to

find the peak power from the given average power.

Solution for (a)

Solving the equation $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$ for the peak voltage V_0 and substituting the known value for V_{rms} gives

Equation:

$$V_0 = \sqrt{2}V_{\text{rms}} = 1.414(120 \text{ V}) = 170 \text{ V}.$$

Discussion for (a)

This means that the AC voltage swings from 170 V to -170 V and back 60 times every second. An equivalent DC voltage is a constant 120 V.

Solution for (b)

Peak power is peak current times peak voltage. Thus,

Equation:

$$P_0 = I_0 V_0 = 2 \left(\frac{1}{2} I_0 V_0 \right) = 2P_{\text{ave}}.$$

We know the average power is 60.0 W, and so

Equation:

$$P_0 = 2(60.0 \text{ W}) = 120 \text{ W}.$$

Discussion

So the power swings from zero to 120 W one hundred twenty times per second (twice each cycle), and the power averages 60 W.

Why Use AC for Power Distribution?

Most large power-distribution systems are AC. Moreover, the power is transmitted at much higher voltages than the 120-V AC (240 V in most parts of the world) we use in homes and on the job. Economies of scale make it cheaper to build a few very large electric power-generation plants than to build numerous small ones. This necessitates sending power long distances, and it is obviously important that energy losses en route be

minimized. High voltages can be transmitted with much smaller power losses than low voltages, as we shall see. (See [\[link\]](#).) For safety reasons, the voltage at the user is reduced to familiar values. The crucial factor is that it is much easier to increase and decrease AC voltages than DC, so AC is used in most large power distribution systems.



Power is distributed over large distances at high voltage to reduce power loss in the transmission lines. The voltages generated at the power plant are stepped up by passive devices called transformers (see [Transformers](#)) to 330,000 volts (or more in some places worldwide). At the point of use, the transformers reduce the voltage transmitted for safe residential and commercial use.

(Credit: GeorgHH, Wikimedia Commons)

Example:**Power Losses Are Less for High-Voltage Transmission**

(a) What current is needed to transmit 100 MW of power at 200 kV? (b) What is the power dissipated by the transmission lines if they have a resistance of $1.00\ \Omega$? (c) What percentage of the power is lost in the transmission lines?

Strategy

We are given $P_{\text{ave}} = 100\text{ MW}$, $V_{\text{rms}} = 200\text{ kV}$, and the resistance of the lines is $R = 1.00\ \Omega$. Using these givens, we can find the current flowing (from $P = IV$) and then the power dissipated in the lines ($P = I^2 R$), and we take the ratio to the total power transmitted.

Solution

To find the current, we rearrange the relationship $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$ and substitute known values. This gives

Equation:

$$I_{\text{rms}} = \frac{P_{\text{ave}}}{V_{\text{rms}}} = \frac{100 \times 10^6\text{ W}}{200 \times 10^3\text{ V}} = 500\text{ A}.$$

Solution

Knowing the current and given the resistance of the lines, the power dissipated in them is found from $P_{\text{ave}} = I_{\text{rms}}^2 R$. Substituting the known values gives

Equation:

$$P_{\text{ave}} = I_{\text{rms}}^2 R = (500\text{ A})^2 (1.00\ \Omega) = 250\text{ kW}.$$

Solution

The percent loss is the ratio of this lost power to the total or input power, multiplied by 100:

Equation:

$$\% \text{ loss} = \frac{250 \text{ kW}}{100 \text{ MW}} \times 100 = 0.250 \%$$

Discussion

One-fourth of a percent is an acceptable loss. Note that if 100 MW of power had been transmitted at 25 kV, then a current of 4000 A would have been needed. This would result in a power loss in the lines of 16.0 MW, or 16.0% rather than 0.250%. The lower the voltage, the more current is needed, and the greater the power loss in the fixed-resistance transmission lines. Of course, lower-resistance lines can be built, but this requires larger and more expensive wires. If superconducting lines could be economically produced, there would be no loss in the transmission lines at all. But, as we shall see in a later chapter, there is a limit to current in superconductors, too. In short, high voltages are more economical for transmitting power, and AC voltage is much easier to raise and lower, so that AC is used in most large-scale power distribution systems.

It is widely recognized that high voltages pose greater hazards than low voltages. But, in fact, some high voltages, such as those associated with common static electricity, can be harmless. So it is not voltage alone that determines a hazard. It is not so widely recognized that AC shocks are often more harmful than similar DC shocks. Thomas Edison thought that AC shocks were more harmful and set up a DC power-distribution system in New York City in the late 1800s. There were bitter fights, in particular between Edison and George Westinghouse and Nikola Tesla, who were advocating the use of AC in early power-distribution systems. AC has prevailed largely due to transformers and lower power losses with high-voltage transmission.

Note:**PhET Explorations: Generator**

Generate electricity with a bar magnet! Discover the physics behind the phenomena by exploring magnets and how you can use them to make a bulb light.

Generato

r

Section Summary

- Direct current (DC) is the flow of electric current in only one direction. It refers to systems where the source voltage is constant.
- The voltage source of an alternating current (AC) system puts out $V = V_0 \sin 2\pi ft$, where V is the voltage at time t , V_0 is the peak voltage, and f is the frequency in hertz.
- In a simple circuit, $I = V/R$ and AC current is $I = I_0 \sin 2\pi ft$, where I is the current at time t , and $I_0 = V_0/R$ is the peak current.
- The average AC power is $P_{\text{ave}} = \frac{1}{2} I_0 V_0$.
- Average (rms) current I_{rms} and average (rms) voltage V_{rms} are $I_{\text{rms}} = \frac{I_0}{\sqrt{2}}$ and $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$, where rms stands for root mean square.
- Thus, $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$.
- Ohm's law for AC is $I_{\text{rms}} = \frac{V_{\text{rms}}}{R}$.
- Expressions for the average power of an AC circuit are $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$, $P_{\text{ave}} = \frac{V_{\text{rms}}^2}{R}$, and $P_{\text{ave}} = I_{\text{rms}}^2 R$, analogous to the expressions for DC circuits.

Conceptual Questions

Exercise:

Problem:

Give an example of a use of AC power other than in the household. Similarly, give an example of a use of DC power other than that supplied by batteries.

Exercise:**Problem:**

Why do voltage, current, and power go through zero 120 times per second for 60-Hz AC electricity?

Exercise:**Problem:**

You are riding in a train, gazing into the distance through its window. As close objects streak by, you notice that the nearby fluorescent lights make *dashed* streaks. Explain.

Problem Exercises**Exercise:****Problem:**

(a) What is the hot resistance of a 25-W light bulb that runs on 120-V AC? (b) If the bulb's operating temperature is 2700°C , what is its resistance at 2600°C ?

Exercise:**Problem:**

Certain heavy industrial equipment uses AC power that has a peak voltage of 679 V. What is the rms voltage?

Solution:

480 V

Exercise:**Problem:**

A certain circuit breaker trips when the rms current is 15.0 A. What is the corresponding peak current?

Exercise:**Problem:**

Military aircraft use 400-Hz AC power, because it is possible to design lighter-weight equipment at this higher frequency. What is the time for one complete cycle of this power?

Solution:

2.50 ms

Exercise:**Problem:**

A North American tourist takes his 25.0-W, 120-V AC razor to Europe, finds a special adapter, and plugs it into 240 V AC. Assuming constant resistance, what power does the razor consume as it is ruined?

Exercise:**Problem:**

In this problem, you will verify statements made at the end of the power losses for [\[link\]](#). (a) What current is needed to transmit 100 MW of power at a voltage of 25.0 kV? (b) Find the power loss in a $1.00\text{ }\Omega$ transmission line. (c) What percent loss does this represent?

Solution:

(a) 4.00 kA

(b) 16.0 MW

(c) 16.0%

Exercise:

Problem:

A small office-building air conditioner operates on 408-V AC and consumes 50.0 kW. (a) What is its effective resistance? (b) What is the cost of running the air conditioner during a hot summer month when it is on 8.00 h per day for 30 days and electricity costs 9.00 cents/kW · h?

Exercise:**Problem:**

What is the peak power consumption of a 120-V AC microwave oven that draws 10.0 A?

Solution:

2.40 kW

Exercise:**Problem:**

What is the peak current through a 500-W room heater that operates on 120-V AC power?

Exercise:**Problem:**

Two different electrical devices have the same power consumption, but one is meant to be operated on 120-V AC and the other on 240-V AC. (a) What is the ratio of their resistances? (b) What is the ratio of their currents? (c) Assuming its resistance is unaffected, by what factor will the power increase if a 120-V AC device is connected to 240-V AC?

Solution:

(a) 4.0

(b) 0.50

(c) 4.0

Exercise:

Problem:

Nichrome wire is used in some radiative heaters. (a) Find the resistance needed if the average power output is to be 1.00 kW utilizing 120-V AC. (b) What length of Nichrome wire, having a cross-sectional area of 5.00mm^2 , is needed if the operating temperature is 500°C ? (c) What power will it draw when first switched on?

Exercise:

Problem:

Find the time after $t = 0$ when the instantaneous voltage of 60-Hz AC first reaches the following values: (a) $V_0/2$ (b) V_0 (c) 0.

Solution:

(a) 1.39 ms

(b) 4.17 ms

(c) 8.33 ms

Exercise:

Problem:

(a) At what two times in the first period following $t = 0$ does the instantaneous voltage in 60-Hz AC equal V_{rms} ? (b) $-V_{\text{rms}}$?

Glossary

direct current

(DC) the flow of electric charge in only one direction

alternating current

(AC) the flow of electric charge that periodically reverses direction

AC voltage

voltage that fluctuates sinusoidally with time, expressed as $V = V_0 \sin 2\pi ft$, where V is the voltage at time t , V_0 is the peak voltage, and f is the frequency in hertz

AC current

current that fluctuates sinusoidally with time, expressed as $I = I_0 \sin 2\pi ft$, where I is the current at time t , I_0 is the peak current, and f is the frequency in hertz

rms current

the root mean square of the current, $I_{\text{rms}} = I_0/\sqrt{2}$, where I_0 is the peak current, in an AC system

rms voltage

the root mean square of the voltage, $V_{\text{rms}} = V_0/\sqrt{2}$, where V_0 is the peak voltage, in an AC system

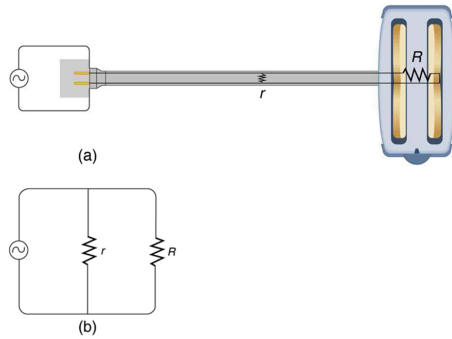
Electric Hazards and the Human Body

- Define thermal hazard, shock hazard, and short circuit.
- Explain what effects various levels of current have on the human body.

There are two known hazards of electricity—thermal and shock. A **thermal hazard** is one where excessive electric power causes undesired thermal effects, such as starting a fire in the wall of a house. A **shock hazard** occurs when electric current passes through a person. Shocks range in severity from painful, but otherwise harmless, to heart-stopping lethality. This section considers these hazards and the various factors affecting them in a quantitative manner. [Electrical Safety: Systems and Devices](#) will consider systems and devices for preventing electrical hazards.

Thermal Hazards

Electric power causes undesired heating effects whenever electric energy is converted to thermal energy at a rate faster than it can be safely dissipated. A classic example of this is the **short circuit**, a low-resistance path between terminals of a voltage source. An example of a short circuit is shown in [\[link\]](#). Insulation on wires leading to an appliance has worn through, allowing the two wires to come into contact. Such an undesired contact with a high voltage is called a *short*. Since the resistance of the short, r , is very small, the power dissipated in the short, $P = V^2/r$, is very large. For example, if V is 120 V and r is 0.100 Ω , then the power is 144 kW, *much* greater than that used by a typical household appliance. Thermal energy delivered at this rate will very quickly raise the temperature of surrounding materials, melting or perhaps igniting them.

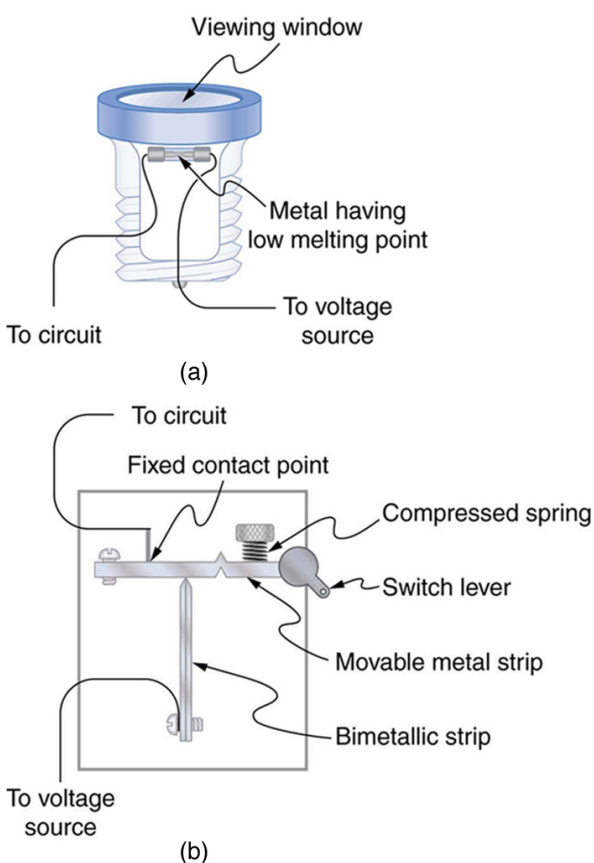


A short circuit is an undesired low-resistance path across a voltage source. (a) Worn insulation on the wires of a toaster allow them to come into contact with a low resistance r . Since $P = V^2/r$, thermal power is created so rapidly that the cord melts or burns. (b) A schematic of the short circuit.

One particularly insidious aspect of a short circuit is that its resistance may actually be decreased due to the increase in temperature. This can happen if the short creates ionization. These charged atoms and molecules are free to move and, thus, lower the resistance r . Since $P = V^2/r$, the power dissipated in the short rises, possibly causing more ionization, more power, and so on. High voltages, such as the 480-V AC used in some industrial applications, lend themselves to this hazard, because higher voltages create higher initial power production in a short.

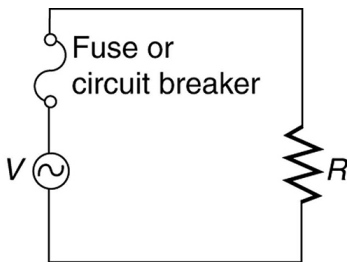
Another serious, but less dramatic, thermal hazard occurs when wires supplying power to a user are overloaded with too great a current. As

discussed in the previous section, the power dissipated in the supply wires is $P = I^2 R_w$, where R_w is the resistance of the wires and I the current flowing through them. If either I or R_w is too large, the wires overheat. For example, a worn appliance cord (with some of its braided wires broken) may have $R_w = 2.00 \, \Omega$ rather than the $0.100 \, \Omega$ it should be. If $10.0 \, \text{A}$ of current passes through the cord, then $P = I^2 R_w = 200 \, \text{W}$ is dissipated in the cord—much more than is safe. Similarly, if a wire with a $0.100 \, \Omega$ resistance is meant to carry a few amps, but is instead carrying $100 \, \text{A}$, it will severely overheat. The power dissipated in the wire will in that case be $P = 1000 \, \text{W}$. Fuses and circuit breakers are used to limit excessive currents. (See [\[link\]](#) and [\[link\]](#).) Each device opens the circuit automatically when a sustained current exceeds safe limits.



(a) A fuse has a metal strip with a low melting point that, when overheated by an excessive

current, permanently breaks the connection of a circuit to a voltage source. (b) A circuit breaker is an automatic but restorable electric switch. The one shown here has a bimetallic strip that bends to the right and into the notch if overheated. The spring then forces the metal strip downward, breaking the electrical connection at the points.



Schematic of a circuit with a fuse or circuit breaker in it.

Fuses and circuit breakers act like automatic switches that open when sustained current exceeds desired limits.

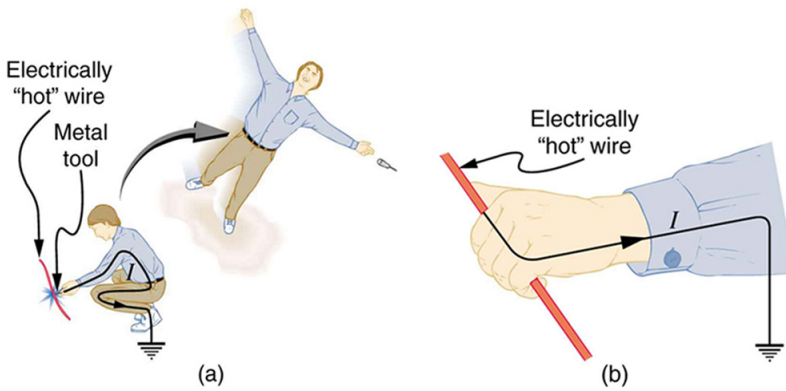
Fuses and circuit breakers for typical household voltages and currents are relatively simple to produce, but those for large voltages and currents experience special problems. For example, when a circuit breaker tries to interrupt the flow of high-voltage electricity, a spark can jump across its points that ionizes the air in the gap and allows the current to continue flowing. Large circuit breakers found in power-distribution systems employ insulating gas and even use jets of gas to blow out such sparks. Here AC is safer than DC, since AC current goes through zero 120 times per second, giving a quick opportunity to extinguish these arcs.

Shock Hazards

Electrical currents through people produce tremendously varied effects. An electrical current can be used to block back pain. The possibility of using electrical current to stimulate muscle action in paralyzed limbs, perhaps allowing paraplegics to walk, is under study. TV dramatizations in which electrical shocks are used to bring a heart attack victim out of ventricular fibrillation (a massively irregular, often fatal, beating of the heart) are more than common. Yet most electrical shock fatalities occur because a current put the heart into fibrillation. A pacemaker uses electrical shocks to stimulate the heart to beat properly. Some fatal shocks do not produce burns, but warts can be safely burned off with electric current (though freezing using liquid nitrogen is now more common). Of course, there are consistent explanations for these disparate effects. The major factors upon which the effects of electrical shock depend are

1. The amount of current I
2. The path taken by the current
3. The duration of the shock
4. The frequency f of the current ($f = 0$ for DC)

[\[link\]](#) gives the effects of electrical shocks as a function of current for a typical accidental shock. The effects are for a shock that passes through the trunk of the body, has a duration of 1 s, and is caused by 60-Hz power.



An electric current can cause muscular contractions with varying effects. (a) The victim is “thrown” backward by involuntary muscle contractions that extend the legs and torso. (b) The victim can’t let go of the wire that is stimulating all the muscles in the hand. Those that close the fingers are stronger than those that open them.

Current (mA)	Effect
1	Threshold of sensation
5	Maximum harmless current
10–20	Onset of sustained muscular contraction; cannot let go for duration of shock; contraction of chest muscles may stop breathing during shock

Current (mA)	Effect
50	Onset of pain
100–300+	Ventricular fibrillation possible; often fatal
300	Onset of burns depending on concentration of current
6000 (6 A)	Onset of sustained ventricular contraction and respiratory paralysis; both cease when shock ends; heartbeat may return to normal; used to defibrillate the heart

Effects of Electrical Shock as a Function of Current^{[footnote](#)}

For an average male shocked through trunk of body for 1 s by 60-Hz AC. Values for females are 60–80% of those listed.

Our bodies are relatively good conductors due to the water in our bodies. Given that larger currents will flow through sections with lower resistance (to be further discussed in the next chapter), electric currents preferentially flow through paths in the human body that have a minimum resistance in a direct path to earth. The earth is a natural electron sink. Wearing insulating shoes, a requirement in many professions, prohibits a pathway for electrons by providing a large resistance in that path. Whenever working with high-power tools (drills), or in risky situations, ensure that you do not provide a pathway for current flow (especially through the heart).

Very small currents pass harmlessly and unfelt through the body. This happens to you regularly without your knowledge. The threshold of sensation is only 1 mA and, although unpleasant, shocks are apparently harmless for currents less than 5 mA. A great number of safety rules take the 5-mA value for the maximum allowed shock. At 10 to 20 mA and above, the current can stimulate sustained muscular contractions much as regular nerve impulses do. People sometimes say they were knocked across the room by a shock, but what really happened was that certain muscles

contracted, propelling them in a manner not of their own choosing. (See [\[link\]](#)(a).) More frightening, and potentially more dangerous, is the “can’t let go” effect illustrated in [\[link\]](#)(b). The muscles that close the fingers are stronger than those that open them, so the hand closes involuntarily on the wire shocking it. This can prolong the shock indefinitely. It can also be a danger to a person trying to rescue the victim, because the rescuer’s hand may close about the victim’s wrist. Usually the best way to help the victim is to give the fist a hard knock/blow/jar with an insulator or to throw an insulator at the fist. Modern electric fences, used in animal enclosures, are now pulsed on and off to allow people who touch them to get free, rendering them less lethal than in the past.

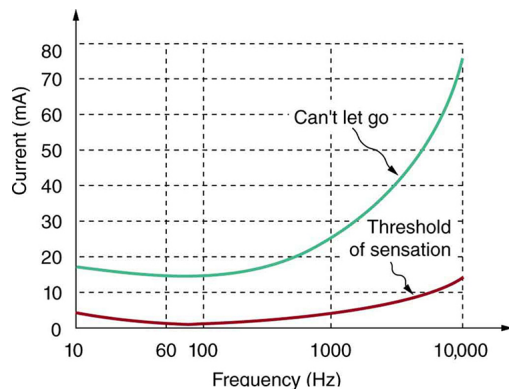
Greater currents may affect the heart. Its electrical patterns can be disrupted, so that it beats irregularly and ineffectively in a condition called “ventricular fibrillation.” This condition often lingers after the shock and is fatal due to a lack of blood circulation. The threshold for ventricular fibrillation is between 100 and 300 mA. At about 300 mA and above, the shock can cause burns, depending on the concentration of current—the more concentrated, the greater the likelihood of burns.

Very large currents cause the heart and diaphragm to contract for the duration of the shock. Both the heart and breathing stop. Interestingly, both often return to normal following the shock. The electrical patterns on the heart are completely erased in a manner that the heart can start afresh with normal beating, as opposed to the permanent disruption caused by smaller currents that can put the heart into ventricular fibrillation. The latter is something like scribbling on a blackboard, whereas the former completely erases it. TV dramatizations of electric shock used to bring a heart attack victim out of ventricular fibrillation also show large paddles. These are used to spread out current passed through the victim to reduce the likelihood of burns.

Current is the major factor determining shock severity (given that other conditions such as path, duration, and frequency are fixed, such as in the table and preceding discussion). A larger voltage is more hazardous, but since $I = V/R$, the severity of the shock depends on the combination of voltage and resistance. For example, a person with dry skin has a resistance

of about $200\text{ k}\Omega$. If he comes into contact with 120-V AC, a current $I = (120\text{ V})/(200\text{ k}\Omega) = 0.6\text{ mA}$ passes harmlessly through him. The same person soaking wet may have a resistance of $10.0\text{ k}\Omega$ and the same 120 V will produce a current of 12 mA—above the “can’t let go” threshold and potentially dangerous.

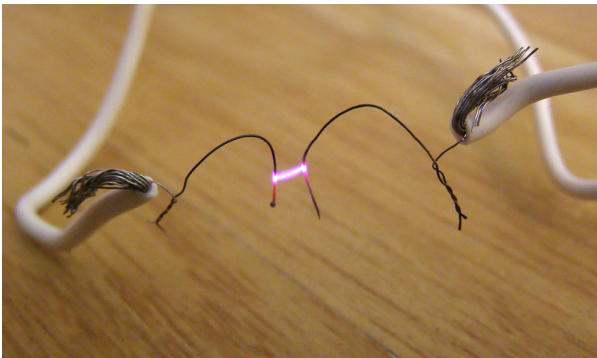
Most of the body’s resistance is in its dry skin. When wet, salts go into ion form, lowering the resistance significantly. The interior of the body has a much lower resistance than dry skin because of all the ionic solutions and fluids it contains. If skin resistance is bypassed, such as by an intravenous infusion, a catheter, or exposed pacemaker leads, a person is rendered **microshock sensitive**. In this condition, currents about 1/1000 those listed in [\[link\]](#) produce similar effects. During open-heart surgery, currents as small as $20\text{ }\mu\text{A}$ can be used to still the heart. Stringent electrical safety requirements in hospitals, particularly in surgery and intensive care, are related to the doubly disadvantaged microshock-sensitive patient. The break in the skin has reduced his resistance, and so the same voltage causes a greater current, and a much smaller current has a greater effect.



Graph of average values
for the threshold of
sensation and the “can’t
let go” current as a
function of frequency.
The lower the value, the

more sensitive the body is
at that frequency.

Factors other than current that affect the severity of a shock are its path, duration, and AC frequency. Path has obvious consequences. For example, the heart is unaffected by an electric shock through the brain, such as may be used to treat manic depression. And it is a general truth that the longer the duration of a shock, the greater its effects. [\[link\]](#) presents a graph that illustrates the effects of frequency on a shock. The curves show the minimum current for two different effects, as a function of frequency. The lower the current needed, the more sensitive the body is at that frequency. Ironically, the body is most sensitive to frequencies near the 50- or 60-Hz frequencies in common use. The body is slightly less sensitive for DC ($f = 0$), mildly confirming Edison's claims that AC presents a greater hazard. At higher and higher frequencies, the body becomes progressively less sensitive to any effects that involve nerves. This is related to the maximum rates at which nerves can fire or be stimulated. At very high frequencies, electrical current travels only on the surface of a person. Thus a wart can be burned off with very high frequency current without causing the heart to stop. (Do not try this at home with 60-Hz AC!) Some of the spectacular demonstrations of electricity, in which high-voltage arcs are passed through the air and over people's bodies, employ high frequencies and low currents. (See [\[link\]](#).) Electrical safety devices and techniques are discussed in detail in [Electrical Safety: Systems and Devices](#).



Is this electric arc dangerous?

The answer depends on the AC frequency and the power involved. (credit: Khimich Alex, Wikimedia Commons)

Section Summary

- The two types of electric hazards are thermal (excessive power) and shock (current through a person).
- Shock severity is determined by current, path, duration, and AC frequency.
- [\[link\]](#) lists shock hazards as a function of current.
- [\[link\]](#) graphs the threshold current for two hazards as a function of frequency.

Conceptual Questions

Exercise:

Problem:

Using an ohmmeter, a student measures the resistance between various points on his body. He finds that the resistance between two points on the same finger is about the same as the resistance between two points on opposite hands—both are several hundred thousand ohms. Furthermore, the resistance decreases when more skin is brought into contact with the probes of the ohmmeter. Finally, there is a dramatic drop in resistance (to a few thousand ohms) when the skin is wet. Explain these observations and their implications regarding skin and internal resistance of the human body.

Exercise:

Problem: What are the two major hazards of electricity?

Exercise:

Problem: Why isn't a short circuit a shock hazard?

Exercise:

Problem:

What determines the severity of a shock? Can you say that a certain voltage is hazardous without further information?

Exercise:

Problem:

An electrified needle is used to burn off warts, with the circuit being completed by having the patient sit on a large butt plate. Why is this plate large?

Exercise:

Problem:

Some surgery is performed with high-voltage electricity passing from a metal scalpel through the tissue being cut. Considering the nature of electric fields at the surface of conductors, why would you expect most of the current to flow from the sharp edge of the scalpel? Do you think high- or low-frequency AC is used?

Exercise:

Problem:

Some devices often used in bathrooms, such as hairdryers, often have safety messages saying "Do not use when the bathtub or basin is full of water." Why is this so?

Exercise:

Problem:

We are often advised to not flick electric switches with wet hands, dry your hand first. We are also advised to never throw water on an electric fire. Why is this so?

Exercise:**Problem:**

Before working on a power transmission line, linemen will touch the line with the back of the hand as a final check that the voltage is zero. Why the back of the hand?

Exercise:**Problem:**

Why is the resistance of wet skin so much smaller than dry, and why do blood and other bodily fluids have low resistances?

Exercise:**Problem:**

Could a person on intravenous infusion (an IV) be microshock sensitive?

Exercise:**Problem:**

In view of the small currents that cause shock hazards and the larger currents that circuit breakers and fuses interrupt, how do they play a role in preventing shock hazards?

Problem Exercises**Exercise:****Problem:**

(a) How much power is dissipated in a short circuit of 240-V AC through a resistance of $0.250\ \Omega$? (b) What current flows?

Solution:

(a) 230 kW

(b) 960 A

Exercise:

Problem:

What voltage is involved in a 1.44-kW short circuit through a $0.100\text{ }\Omega$ resistance?

Exercise:

Problem:

Find the current through a person and identify the likely effect on her if she touches a 120-V AC source: (a) if she is standing on a rubber mat and offers a total resistance of $300\text{ k}\Omega$; (b) if she is standing barefoot on wet grass and has a resistance of only $4000\text{ k}\Omega$.

Solution:

(a) 0.400 mA, no effect

(b) 26.7 mA, muscular contraction for duration of the shock (can't let go)

Exercise:

Problem:

While taking a bath, a person touches the metal case of a radio. The path through the person to the drainpipe and ground has a resistance of $4000\text{ }\Omega$. What is the smallest voltage on the case of the radio that could cause ventricular fibrillation?

Exercise:

Problem:

Foolishly trying to fish a burning piece of bread from a toaster with a metal butter knife, a man comes into contact with 120-V AC. He does not even feel it since, luckily, he is wearing rubber-soled shoes. What is the minimum resistance of the path the current follows through the person?

Solution:

$$1.20 \times 10^5 \, \Omega$$

Exercise:**Problem:**

(a) During surgery, a current as small as $20.0 \, \mu\text{A}$ applied directly to the heart may cause ventricular fibrillation. If the resistance of the exposed heart is $300 \, \Omega$, what is the smallest voltage that poses this danger? (b) Does your answer imply that special electrical safety precautions are needed?

Exercise:**Problem:**

(a) What is the resistance of a 220-V AC short circuit that generates a peak power of 96.8 kW? (b) What would the average power be if the voltage was 120 V AC?

Solution:

(a) $1.00 \, \Omega$

(b) 14.4 kW

Exercise:

Problem:

A heart defibrillator passes 10.0 A through a patient's torso for 5.00 ms in an attempt to restore normal beating. (a) How much charge passed? (b) What voltage was applied if 500 J of energy was dissipated? (c) What was the path's resistance? (d) Find the temperature increase caused in the 8.00 kg of affected tissue.

Exercise:**Problem: Integrated Concepts**

A short circuit in a 120-V appliance cord has a $0.500\text{-}\Omega$ resistance. Calculate the temperature rise of the 2.00 g of surrounding materials, assuming their specific heat capacity is $0.200\text{ cal/g}\cdot^{\circ}\text{C}$ and that it takes 0.0500 s for a circuit breaker to interrupt the current. Is this likely to be damaging?

Solution:

Temperature increases 860°C . It is very likely to be damaging.

Exercise:**Problem: Construct Your Own Problem**

Consider a person working in an environment where electric currents might pass through her body. Construct a problem in which you calculate the resistance of insulation needed to protect the person from harm. Among the things to be considered are the voltage to which the person might be exposed, likely body resistance (dry, wet, ...), and acceptable currents (safe but sensed, safe and unfelt, ...).

Glossary

thermal hazard

a hazard in which electric current causes undesired thermal effects

shock hazard

when electric current passes through a person

short circuit

also known as a “short,” a low-resistance path between terminals of a voltage source

microshock sensitive

a condition in which a person’s skin resistance is bypassed, possibly by a medical procedure, rendering the person vulnerable to electrical shock at currents about 1/1000 the normally required level

Introduction to Magnetism

class="introduction"

The
magnificent
spectacle
of the
Aurora
Borealis, or
northern
lights,
glows in
the
northern
sky above
Bear Lake
near
Eielson Air
Force Base,
Alaska.
Shaped by
the Earth's
magnetic
field, this
light is
produced
by
radiation
spewed
from solar
storms.
(credit:
Senior
Airman
Joshua
Strang, via
Flickr)



One evening, an Alaskan sticks a note to his refrigerator with a small magnet. Through the kitchen window, the Aurora Borealis glows in the night sky. This grand spectacle is shaped by the same force that holds the note to the refrigerator.

People have been aware of magnets and magnetism for thousands of years. The earliest records date to well before the time of Christ, particularly in a region of Asia Minor called Magnesia (the name of this region is the source of words like *magnetic*). Magnetic rocks found in Magnesia, which is now part of western Turkey, stimulated interest during ancient times. A practical application for magnets was found later, when they were employed as navigational compasses. The use of magnets in compasses resulted not only in improved long-distance sailing, but also in the names of “north” and “south” being given to the two types of magnetic poles.

Today magnetism plays many important roles in our lives. Physicists’ understanding of magnetism has enabled the development of technologies that affect our everyday lives. The iPod in your purse or backpack, for example, wouldn’t have been possible without the applications of magnetism and electricity on a small scale.

The discovery that weak changes in a magnetic field in a thin film of iron and chromium could bring about much larger changes in electrical resistance was one of the first large successes of nanotechnology. The 2007 Nobel Prize in Physics went to Albert Fert from France and Peter Grunberg from Germany for this discovery of *giant magnetoresistance* and its applications to computer memory.

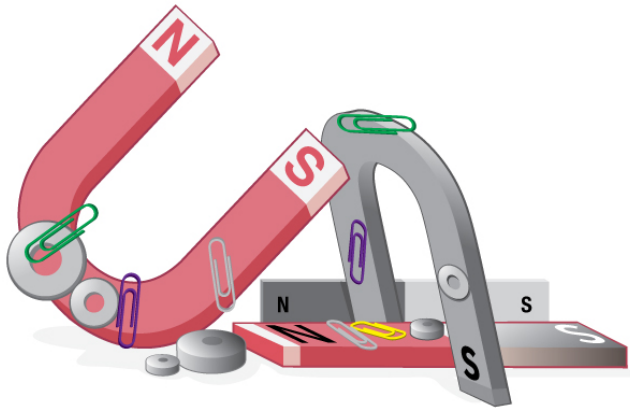
All electric motors, with uses as diverse as powering refrigerators, starting cars, and moving elevators, contain magnets. Generators, whether producing hydroelectric power or running bicycle lights, use magnetic fields. Recycling facilities employ magnets to separate iron from other refuse. Hundreds of millions of dollars are spent annually on magnetic containment of fusion as a future energy source. Magnetic resonance imaging (MRI) has become an important diagnostic tool in the field of medicine, and the use of magnetism to explore brain activity is a subject of contemporary research and development. The list of applications also includes computer hard drives, tape recording, detection of inhaled asbestos, and levitation of high-speed trains. Magnetism is used to explain atomic energy levels, cosmic rays, and charged particles trapped in the Van Allen belts. Once again, we will find all these disparate phenomena are linked by a small number of underlying physical principles.



Engineering of
technology like iPods
would not be possible
without a deep
understanding
magnetism. (credit: Jesse!
S?, Flickr)

Magnets

- Describe the difference between the north and south poles of a magnet.
- Describe how magnetic poles interact with each other.



Magnets come in various shapes, sizes, and strengths. All have both a north pole and a south pole. There is never an isolated pole (a monopole).

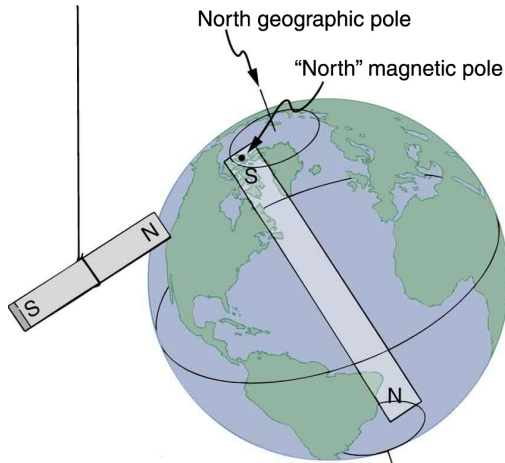
All magnets attract iron, such as that in a refrigerator door. However, magnets may attract or repel other magnets. Experimentation shows that all magnets have two poles. If freely suspended, one pole will point toward the north. The two poles are thus named the **north magnetic pole** and the **south magnetic pole** (or more properly, north-seeking and south-seeking poles, for the attractions in those directions).

Note:

Universal Characteristics of Magnets and Magnetic Poles

It is a universal characteristic of all magnets that *like poles repel and unlike poles attract*. (Note the similarity with electrostatics: unlike charges attract and like charges repel.)

Further experimentation shows that it is *impossible to separate north and south poles* in the manner that + and – charges can be separated.



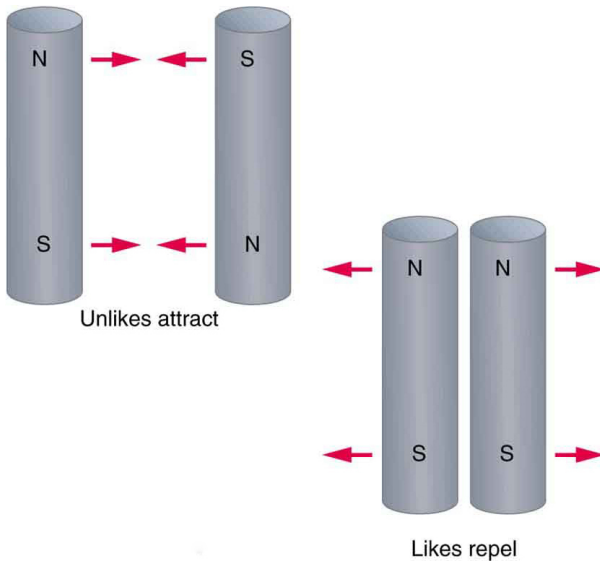
One end of a bar magnet is suspended from a thread that points toward north. The magnet's two poles are labeled N and S for north-seeking and south-seeking poles, respectively.

Note:

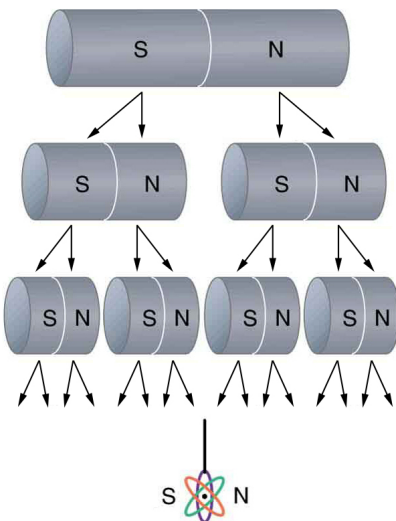
Misconception Alert: Earth's Geographic North Pole Hides an S

The Earth acts like a very large bar magnet with its south-seeking pole near the geographic North Pole. That is why the north pole of your compass is attracted toward the geographic north pole of the Earth—because the magnetic pole that is near the geographic North Pole is actually a south magnetic pole! Confusion arises because the geographic term “North Pole” has come to be used (incorrectly) for the magnetic pole that is near the

North Pole. Thus, “North magnetic pole” is actually a misnomer—it should be called the South magnetic pole.



Unlike poles attract, whereas
like poles repel.



North and south
poles always occur
in pairs. Attempts

to separate them
result in more pairs
of poles. If we
continue to split the
magnet, we will
eventually get
down to an iron
atom with a north
pole and a south
pole—these, too,
cannot be
separated.

The fact that magnetic poles always occur in pairs of north and south is true from the very large scale—for example, sunspots always occur in pairs that are north and south magnetic poles—all the way down to the very small scale. Magnetic atoms have both a north pole and a south pole, as do many types of subatomic particles, such as electrons, protons, and neutrons.

Note:

Making Connections: Take-Home Experiment—Refrigerator Magnets

We know that like magnetic poles repel and unlike poles attract. See if you can show this for two refrigerator magnets. Will the magnets stick if you turn them over? Why do they stick to the door anyway? What can you say about the magnetic properties of the door next to the magnet? Do refrigerator magnets stick to metal or plastic spoons? Do they stick to all types of metal?

Section Summary

- Magnetism is a subject that includes the properties of magnets, the effect of the magnetic force on moving charges and currents, and the

- creation of magnetic fields by currents.
- There are two types of magnetic poles, called the north magnetic pole and south magnetic pole.
 - North magnetic poles are those that are attracted toward the Earth's geographic north pole.
 - Like poles repel and unlike poles attract.
 - Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.

Conceptual Questions

Exercise:

Problem:

Volcanic and other such activity at the mid-Atlantic ridge extrudes material to fill the gap between separating tectonic plates associated with continental drift. The magnetization of rocks is found to reverse in a coordinated manner with distance from the ridge. What does this imply about the Earth's magnetic field and how could the knowledge of the spreading rate be used to give its historical record?

Glossary

north magnetic pole

the end or the side of a magnet that is attracted toward Earth's geographic north pole

south magnetic pole

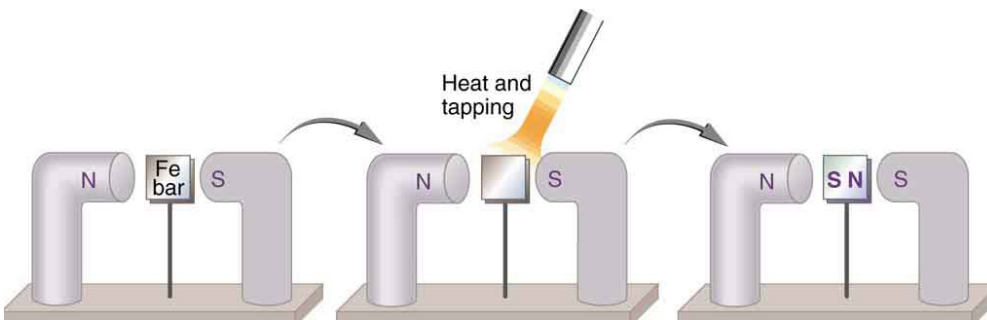
the end or the side of a magnet that is attracted toward Earth's geographic south pole

Ferromagnets and Electromagnets

- Define ferromagnet.
- Describe the role of magnetic domains in magnetization.
- Explain the significance of the Curie temperature.
- Describe the relationship between electricity and magnetism.

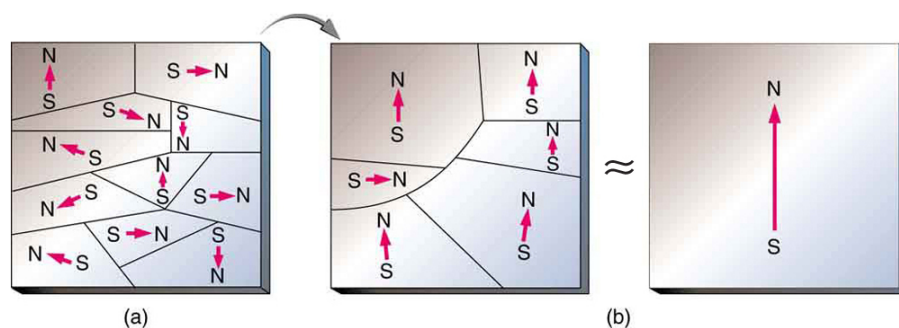
Ferromagnets

Only certain materials, such as iron, cobalt, nickel, and gadolinium, exhibit strong magnetic effects. Such materials are called **ferromagnetic**, after the Latin word for iron, *ferrum*. A group of materials made from the alloys of the rare earth elements are also used as strong and permanent magnets; a popular one is neodymium. Other materials exhibit weak magnetic effects, which are detectable only with sensitive instruments. Not only do ferromagnetic materials respond strongly to magnets (the way iron is attracted to magnets), they can also be **magnetized** themselves—that is, they can be induced to be magnetic or made into permanent magnets.



An unmagnetized piece of iron is placed between two magnets, heated, and then cooled, or simply tapped when cold. The iron becomes a permanent magnet with the poles aligned as shown: its south pole is adjacent to the north pole of the original magnet, and its north pole is adjacent to the south pole of the original magnet. Note that there are attractive forces between the magnets.

When a magnet is brought near a previously unmagnetized ferromagnetic material, it causes local magnetization of the material with unlike poles closest, as in [\[link\]](#). (This results in the attraction of the previously unmagnetized material to the magnet.) What happens on a microscopic scale is illustrated in [\[link\]](#). The regions within the material called **domains** act like small bar magnets. Within domains, the poles of individual atoms are aligned. Each atom acts like a tiny bar magnet. Domains are small and randomly oriented in an unmagnetized ferromagnetic object. In response to an external magnetic field, the domains may grow to millimeter size, aligning themselves as shown in [\[link\]](#)(b). This induced magnetization can be made permanent if the material is heated and then cooled, or simply tapped in the presence of other magnets.



(a) An unmagnetized piece of iron (or other ferromagnetic material) has randomly oriented domains. (b) When magnetized by an external field, the domains show greater alignment, and some grow at the expense of others. Individual atoms are aligned within domains; each atom acts like a tiny bar magnet.

Conversely, a permanent magnet can be demagnetized by hard blows or by heating it in the absence of another magnet. Increased thermal motion at higher temperature can disrupt and randomize the orientation and the size of

the domains. There is a well-defined temperature for ferromagnetic materials, which is called the **Curie temperature**, above which they cannot be magnetized. The Curie temperature for iron is 1043 K (770°C), which is well above room temperature. There are several elements and alloys that have Curie temperatures much lower than room temperature and are ferromagnetic only below those temperatures.

Electromagnets

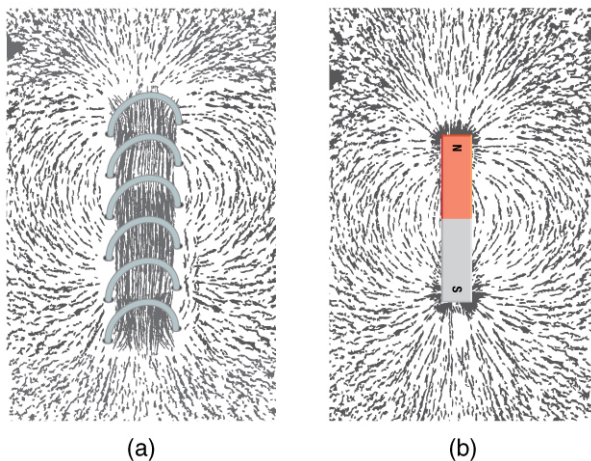
Early in the 19th century, it was discovered that electrical currents cause magnetic effects. The first significant observation was by the Danish scientist Hans Christian Oersted (1777–1851), who found that a compass needle was deflected by a current-carrying wire. This was the first significant evidence that the movement of charges had any connection with magnets. **Electromagnetism** is the use of electric current to make magnets. These temporarily induced magnets are called **electromagnets**. Electromagnets are employed for everything from a wrecking yard crane that lifts scrapped cars to controlling the beam of a 90-km-circumference particle accelerator to the magnets in medical imaging machines (See [\[link\]](#)).



Instrument for magnetic resonance imaging (MRI). The device uses a superconducting

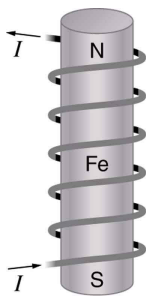
cylindrical coil for the main magnetic field. The patient goes into this “tunnel” on the gurney.
(credit: Bill McChesney, Flickr)

[\[link\]](#) shows that the response of iron filings to a current-carrying coil and to a permanent bar magnet. The patterns are similar. In fact, electromagnets and ferromagnets have the same basic characteristics—for example, they have north and south poles that cannot be separated and for which like poles repel and unlike poles attract.



Iron filings near (a) a current-carrying coil and (b) a magnet act like tiny compass needles, showing the shape of their fields. Their response to a current-carrying coil and a permanent magnet is seen to be very similar, especially near the ends of the coil and the magnet.

Combining a ferromagnet with an electromagnet can produce particularly strong magnetic effects. (See [\[link\]](#).) Whenever strong magnetic effects are needed, such as lifting scrap metal, or in particle accelerators, electromagnets are enhanced by ferromagnetic materials. Limits to how strong the magnets can be made are imposed by coil resistance (it will overheat and melt at sufficiently high current), and so superconducting magnets may be employed. These are still limited, because superconducting properties are destroyed by too great a magnetic field.

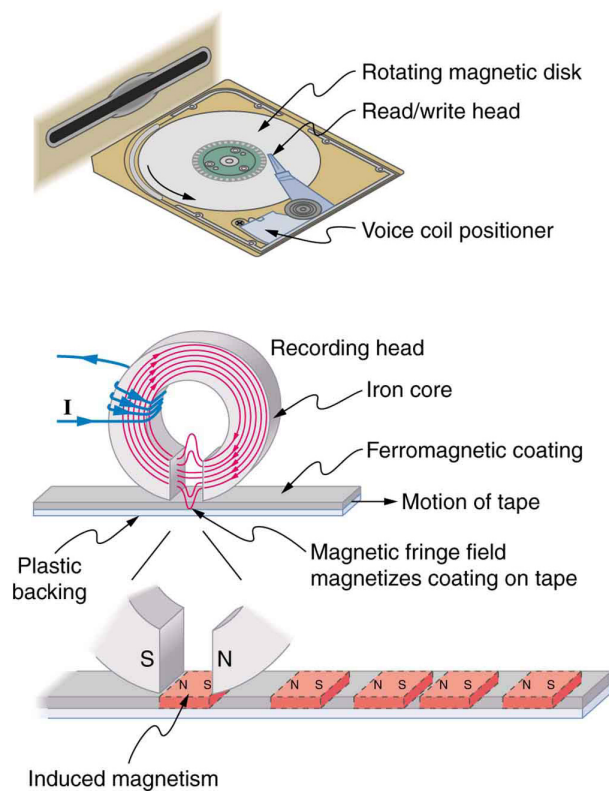


An
electromagnet
with a
ferromagnetic
core can
produce very
strong
magnetic
effects.

Alignment of
domains in the
core produces
a magnet, the
poles of which
are aligned
with the
electromagnet

.

[\[link\]](#) shows a few uses of combinations of electromagnets and ferromagnets. Ferromagnetic materials can act as memory devices, because the orientation of the magnetic fields of small domains can be reversed or erased. Magnetic information storage on videotapes and computer hard drives are among the most common applications. This property is vital in our digital world.



An electromagnet induces regions of permanent magnetism on a floppy disk coated with a ferromagnetic material. The information stored here is digital (a region is either magnetic or not); in other applications, it can be analog

(with a varying strength), such
as on audiotapes.

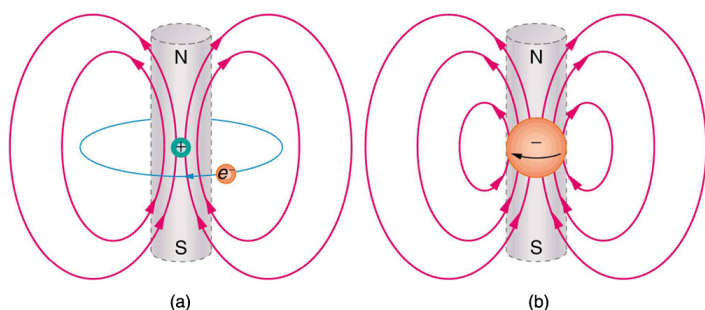
Current: The Source of All Magnetism

An electromagnet creates magnetism with an electric current. In later sections we explore this more quantitatively, finding the strength and direction of magnetic fields created by various currents. But what about ferromagnets? [\[link\]](#) shows models of how electric currents create magnetism at the submicroscopic level. (Note that we cannot directly observe the paths of individual electrons about atoms, and so a model or visual image, consistent with all direct observations, is made. We can directly observe the electron's orbital angular momentum, its spin momentum, and subsequent magnetic moments, all of which are explained with electric-current-creating subatomic magnetism.) Currents, including those associated with other submicroscopic particles like protons, allow us to explain ferromagnetism and all other magnetic effects. Ferromagnetism, for example, results from an internal cooperative alignment of electron spins, possible in some materials but not in others.

Crucial to the statement that electric current is the source of all magnetism is the fact that it is impossible to separate north and south magnetic poles. (This is far different from the case of positive and negative charges, which are easily separated.) A current loop always produces a magnetic dipole—that is, a magnetic field that acts like a north pole and south pole pair. Since isolated north and south magnetic poles, called **magnetic monopoles**, are not observed, currents are used to explain all magnetic effects. If magnetic monopoles did exist, then we would have to modify this underlying connection that all magnetism is due to electrical current. There is no known reason that magnetic monopoles should not exist—they are simply never observed—and so searches at the subnuclear level continue. If they *do not* exist, we would like to find out why not. If they *do* exist, we would like to see evidence of them.

Note:**Electric Currents and Magnetism**

Electric current is the source of all magnetism.



(a) In the planetary model of the atom, an electron orbits a nucleus, forming a closed-current loop and producing a magnetic field with a north pole and a south pole. (b) Electrons have spin and can be crudely pictured as rotating charge, forming a current that produces a magnetic field with a north pole and a south pole. Neither the planetary model nor the image of a spinning electron is completely consistent with modern physics. However, they do provide a useful way of understanding phenomena.

Note:**PhET Explorations: Magnets and Electromagnets**

Explore the interactions between a compass and bar magnet. Discover how you can use a battery and wire to make a magnet! Can you make it a stronger magnet? Can you make the magnetic field reverse?

Section Summary

- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.
- All magnetism is created by electric current.
- Ferromagnetic materials, such as iron, are those that exhibit strong magnetic effects.
- The atoms in ferromagnetic materials act like small magnets (due to currents within the atoms) and can be aligned, usually in millimeter-sized regions called domains.
- Domains can grow and align on a larger scale, producing permanent magnets. Such a material is magnetized, or induced to be magnetic.
- Above a material's Curie temperature, thermal agitation destroys the alignment of atoms, and ferromagnetism disappears.
- Electromagnets employ electric currents to make magnetic fields, often aided by induced fields in ferromagnetic materials.

Glossary

ferromagnetic

materials, such as iron, cobalt, nickel, and gadolinium, that exhibit strong magnetic effects

magnetized

to be turned into a magnet; to be induced to be magnetic

domains

regions within a material that behave like small bar magnets

Curie temperature

the temperature above which a ferromagnetic material cannot be magnetized

electromagnetism

the use of electrical currents to induce magnetism

electromagnet

an object that is temporarily magnetic when an electrical current is passed through it

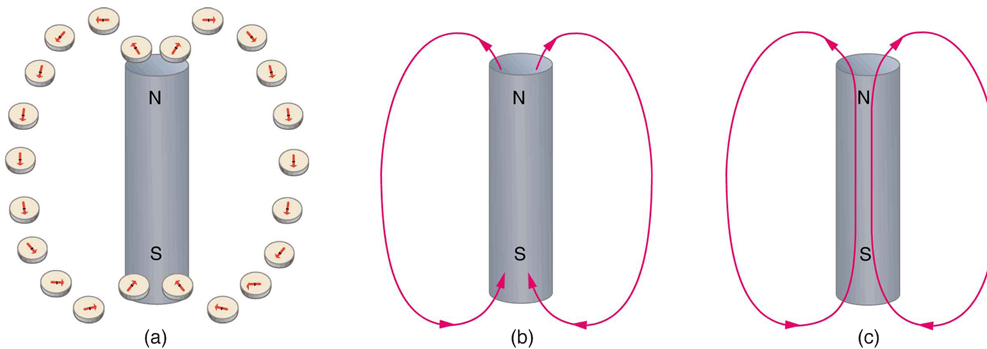
magnetic monopoles

an isolated magnetic pole; a south pole without a north pole, or vice versa (no magnetic monopole has ever been observed)

Magnetic Fields and Magnetic Field Lines

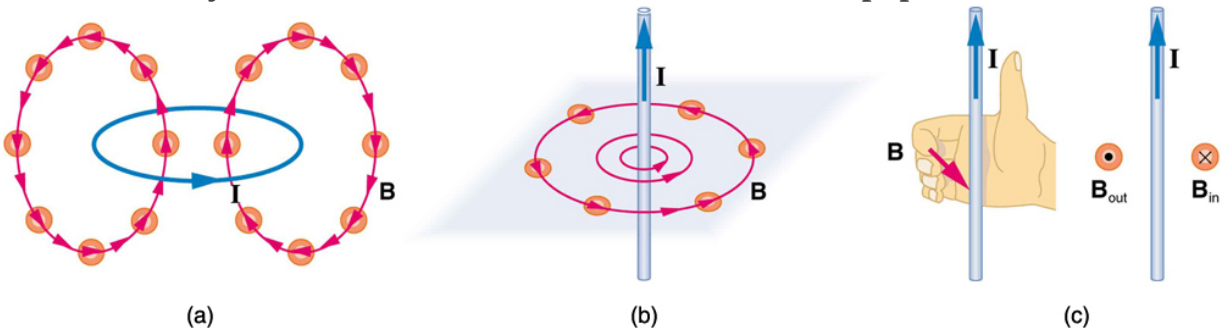
- Define magnetic field and describe the magnetic field lines of various magnetic fields.

Einstein is said to have been fascinated by a compass as a child, perhaps musing on how the needle felt a force without direct physical contact. His ability to think deeply and clearly about action at a distance, particularly for gravitational, electric, and magnetic forces, later enabled him to create his revolutionary theory of relativity. Since magnetic forces act at a distance, we define a **magnetic field** to represent magnetic forces. The pictorial representation of **magnetic field lines** is very useful in visualizing the strength and direction of the magnetic field. As shown in [\[link\]](#), the **direction of magnetic field lines** is defined to be the direction in which the north end of a compass needle points. The magnetic field is traditionally called the ***B*-field**.



Magnetic field lines are defined to have the direction that a small compass points when placed at a location. (a) If small compasses are used to map the magnetic field around a bar magnet, they will point in the directions shown: away from the north pole of the magnet, toward the south pole of the magnet. (Recall that the Earth's north magnetic pole is really a south pole in terms of definitions of poles on a bar magnet.) (b) Connecting the arrows gives continuous magnetic field lines. The strength of the field is proportional to the closeness (or density) of the lines. (c) If the interior of the magnet could be probed, the field lines would be found to form continuous closed loops.

Small compasses used to test a magnetic field will not disturb it. (This is analogous to the way we tested electric fields with a small test charge. In both cases, the fields represent only the object creating them and not the probe testing them.) [\[link\]](#) shows how the magnetic field appears for a current loop and a long straight wire, as could be explored with small compasses. A small compass placed in these fields will align itself parallel to the field line at its location, with its north pole pointing in the direction of B . Note the symbols used for field into and out of the paper.



Small compasses could be used to map the fields shown here. (a) The magnetic field of a circular current loop is similar to that of a bar magnet. (b) A long and straight wire creates a field with magnetic field lines forming circular loops. (c) When the wire is in the plane of the paper, the field is perpendicular to the paper. Note that the symbols used for the field pointing inward (like the tail of an arrow) and the field pointing outward (like the tip of an arrow).

Note:

Making Connections: Concept of a Field

A field is a way of mapping forces surrounding any object that can act on another object at a distance without apparent physical connection. The field represents the object generating it. Gravitational fields map

gravitational forces, electric fields map electrical forces, and magnetic fields map magnetic forces.

Extensive exploration of magnetic fields has revealed a number of hard-and-fast rules. We use magnetic field lines to represent the field (the lines are a pictorial tool, not a physical entity in and of themselves). The properties of magnetic field lines can be summarized by these rules:

1. The direction of the magnetic field is tangent to the field line at any point in space. A small compass will point in the direction of the field line.
2. The strength of the field is proportional to the closeness of the lines. It is exactly proportional to the number of lines per unit area perpendicular to the lines (called the areal density).
3. Magnetic field lines can never cross, meaning that the field is unique at any point in space.
4. Magnetic field lines are continuous, forming closed loops without beginning or end. They go from the north pole to the south pole.

The last property is related to the fact that the north and south poles cannot be separated. It is a distinct difference from electric field lines, which begin and end on the positive and negative charges. If magnetic monopoles existed, then magnetic field lines would begin and end on them.

Section Summary

- Magnetic fields can be pictorially represented by magnetic field lines, the properties of which are as follows:
 1. The field is tangent to the magnetic field line.
 2. Field strength is proportional to the line density.
 3. Field lines cannot cross.
 4. Field lines are continuous loops.

Conceptual Questions

Exercise:**Problem:**

Explain why the magnetic field would not be unique (that is, not have a single value) at a point in space where magnetic field lines might cross. (Consider the direction of the field at such a point.)

Exercise:**Problem:**

List the ways in which magnetic field lines and electric field lines are similar. For example, the field direction is tangent to the line at any point in space. Also list the ways in which they differ. For example, electric force is parallel to electric field lines, whereas magnetic force on moving charges is perpendicular to magnetic field lines.

Exercise:**Problem:**

Noting that the magnetic field lines of a bar magnet resemble the electric field lines of a pair of equal and opposite charges, do you expect the magnetic field to rapidly decrease in strength with distance from the magnet? Is this consistent with your experience with magnets?

Exercise:**Problem:**

Is the Earth's magnetic field parallel to the ground at all locations? If not, where is it parallel to the surface? Is its strength the same at all locations? If not, where is it greatest?

Glossary

magnetic field

the representation of magnetic forces

B-field

another term for magnetic field

magnetic field lines

the pictorial representation of the strength and the direction of a magnetic field

direction of magnetic field lines

the direction that the north end of a compass needle points

Magnetic Field Strength: Force on a Moving Charge in a Magnetic Field

- Describe the effects of magnetic fields on moving charges.
- Use the right hand rule 1 to determine the velocity of a charge, the direction of the magnetic field, and the direction of the magnetic force on a moving charge.
- Calculate the magnetic force on a moving charge.

What is the mechanism by which one magnet exerts a force on another?

The answer is related to the fact that all magnetism is caused by current, the flow of charge. *Magnetic fields exert forces on moving charges*, and so they exert forces on other magnets, all of which have moving charges.

Right Hand Rule 1

The magnetic force on a moving charge is one of the most fundamental known. Magnetic force is as important as the electrostatic or Coulomb force. Yet the magnetic force is more complex, in both the number of factors that affects it and in its direction, than the relatively simple Coulomb force. The magnitude of the **magnetic force** F on a charge q moving at a speed v in a magnetic field of strength B is given by

Equation:

$$F = qvB \sin \theta,$$

where θ is the angle between the directions of \mathbf{v} and \mathbf{B} . This force is often called the **Lorentz force**. In fact, this is how we define the magnetic field strength B —in terms of the force on a charged particle moving in a magnetic field. The SI unit for magnetic field strength B is called the **tesla** (T) after the eccentric but brilliant inventor Nikola Tesla (1856–1943). To determine how the tesla relates to other SI units, we solve $F = qvB \sin \theta$ for B .

Equation:

$$B = \frac{F}{qv \sin \theta}$$

Because $\sin \theta$ is unitless, the tesla is

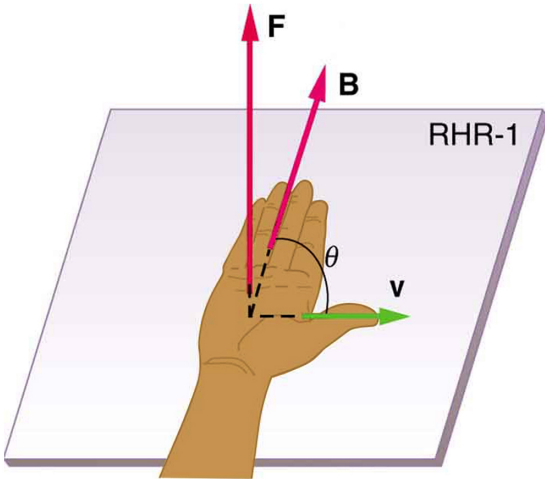
Equation:

$$1 \text{ T} = \frac{1 \text{ N}}{\text{C} \cdot \text{m/s}} = \frac{1 \text{ N}}{\text{A} \cdot \text{m}}$$

(note that $\text{C/s} = \text{A}$).

Another smaller unit, called the **gauss** (G), where $1 \text{ G} = 10^{-4} \text{ T}$, is sometimes used. The strongest permanent magnets have fields near 2 T; superconducting electromagnets may attain 10 T or more. The Earth's magnetic field on its surface is only about $5 \times 10^{-5} \text{ T}$, or 0.5 G.

The *direction* of the magnetic force **F** is perpendicular to the plane formed by **v** and **B**, as determined by the **right hand rule 1** (or RHR-1), which is illustrated in [\[link\]](#). RHR-1 states that, to determine the direction of the magnetic force on a positive moving charge, you point the thumb of the right hand in the direction of **v**, the fingers in the direction of **B**, and a perpendicular to the palm points in the direction of **F**. One way to remember this is that there is one velocity, and so the thumb represents it. There are many field lines, and so the fingers represent them. The force is in the direction you would push with your palm. The force on a negative charge is in exactly the opposite direction to that on a positive charge.



$$F = qvB \sin \theta$$

$\mathbf{F} \perp \text{plane of } \mathbf{v} \text{ and } \mathbf{B}$

Magnetic fields exert forces on moving charges. This force is one of the most basic known. The direction of the magnetic force on a moving charge is perpendicular to the plane formed by \mathbf{v} and \mathbf{B} and follows right hand rule—1 (RHR-1) as shown. The magnitude of the force is proportional to q , v , B , and the sine of the angle between \mathbf{v} and \mathbf{B} .

Note:

Making Connections: Charges and Magnets

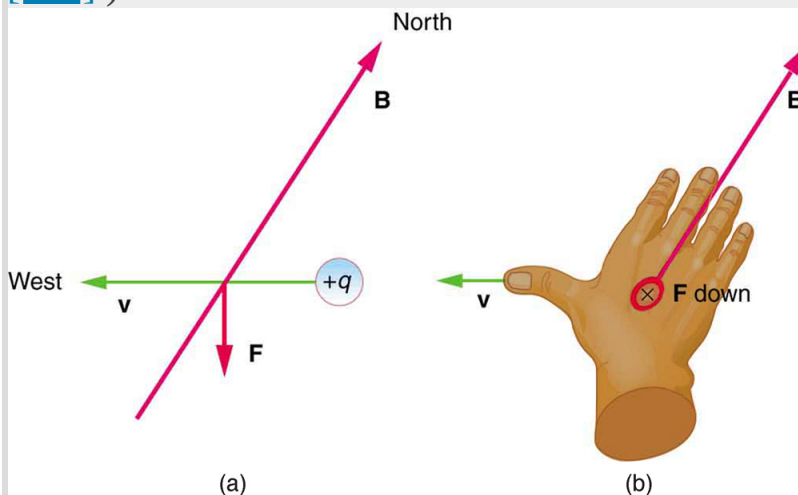
There is no magnetic force on static charges. However, there is a magnetic force on moving charges. When charges are stationary, their electric fields do not affect magnets. But, when charges move, they produce magnetic

fields that exert forces on other magnets. When there is relative motion, a connection between electric and magnetic fields emerges—each affects the other.

Example:

Calculating Magnetic Force: Earth's Magnetic Field on a Charged Glass Rod

With the exception of compasses, you seldom see or personally experience forces due to the Earth's small magnetic field. To illustrate this, suppose that in a physics lab you rub a glass rod with silk, placing a 20-nC positive charge on it. Calculate the force on the rod due to the Earth's magnetic field, if you throw it with a horizontal velocity of 10 m/s due west in a place where the Earth's field is due north parallel to the ground. (The direction of the force is determined with right hand rule 1 as shown in [\[link\]](#).)



A positively charged object moving due west in a region where the Earth's magnetic field is due north experiences a force that is straight down as shown. A negative charge moving in the same direction would feel a force straight up.

Strategy

We are given the charge, its velocity, and the magnetic field strength and direction. We can thus use the equation $F = qvB \sin \theta$ to find the force.

Solution

The magnetic force is

Equation:

$$F = qvb \sin \theta.$$

We see that $\sin \theta = 1$, since the angle between the velocity and the direction of the field is 90° . Entering the other given quantities yields

Equation:

$$\begin{aligned} F &= (20 \times 10^{-9} \text{ C})(10 \text{ m/s})(5 \times 10^{-5} \text{ T}) \\ &= 1 \times 10^{-11} (\text{C} \cdot \text{m/s}) \left(\frac{\text{N}}{\text{C} \cdot \text{m/s}} \right) = 1 \times 10^{-11} \text{ N}. \end{aligned}$$

Discussion

This force is completely negligible on any macroscopic object, consistent with experience. (It is calculated to only one digit, since the Earth's field varies with location and is given to only one digit.) The Earth's magnetic field, however, does produce very important effects, particularly on submicroscopic particles. Some of these are explored in [Force on a Moving Charge in a Magnetic Field: Examples and Applications](#).

Section Summary

- Magnetic fields exert a force on a moving charge q , the magnitude of which is

Equation:

$$F = qvB \sin \theta,$$

where θ is the angle between the directions of v and B .

- The SI unit for magnetic field strength B is the tesla (T), which is related to other units by

Equation:

$$1 \text{ T} = \frac{1 \text{ N}}{\text{C} \cdot \text{m/s}} = \frac{1 \text{ N}}{\text{A} \cdot \text{m}}.$$

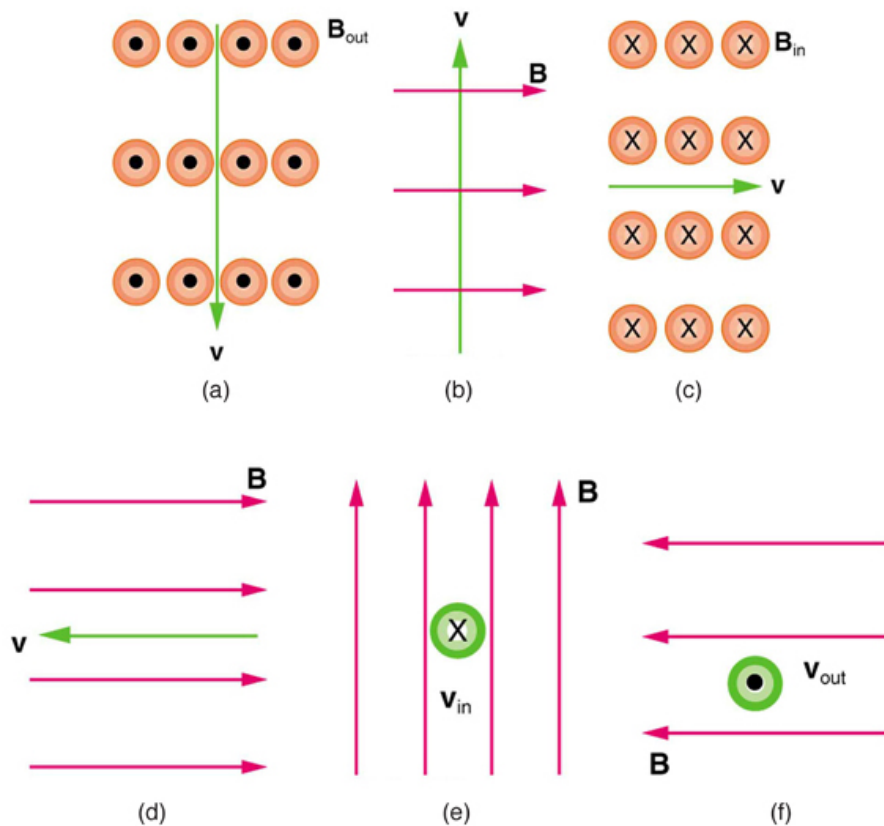
- The *direction* of the force on a moving charge is given by right hand rule 1 (RHR-1): Point the thumb of the right hand in the direction of v , the fingers in the direction of B , and a perpendicular to the palm points in the direction of F .
- The force is perpendicular to the plane formed by \mathbf{v} and \mathbf{B} . Since the force is zero if \mathbf{v} is parallel to \mathbf{B} , charged particles often follow magnetic field lines rather than cross them.

Conceptual Questions**Exercise:****Problem:**

If a charged particle moves in a straight line through some region of space, can you say that the magnetic field in that region is necessarily zero?

Problems & Exercises**Exercise:****Problem:**

What is the direction of the magnetic force on a positive charge that moves as shown in each of the six cases shown in [\[link\]](#)?



Solution:

- (a) Left (West)
- (b) Into the page
- (c) Up (North)
- (d) No force
- (e) Right (East)
- (f) Down (South)

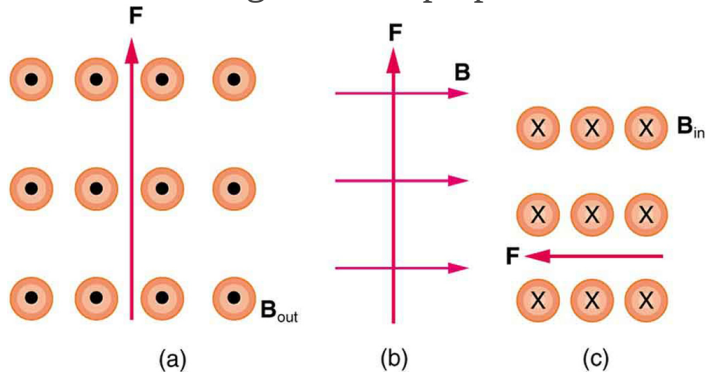
Exercise:

Problem: Repeat [\[link\]](#) for a negative charge.

Exercise:

Problem:

What is the direction of the velocity of a negative charge that experiences the magnetic force shown in each of the three cases in [\[link\]](#), assuming it moves perpendicular to \mathbf{B} ?

**Solution:**

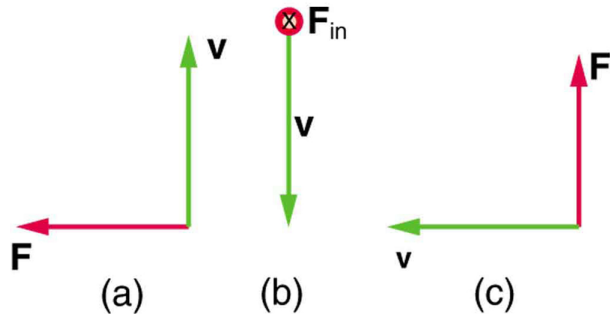
- (a) East (right)
- (b) Into page
- (c) South (down)

Exercise:

Problem: Repeat [\[link\]](#) for a positive charge.

Exercise:**Problem:**

What is the direction of the magnetic field that produces the magnetic force on a positive charge as shown in each of the three cases in the figure below, assuming \mathbf{B} is perpendicular to \mathbf{v} ?



Solution:

(a) Into page

(b) West (left)

(c) Out of page

Exercise:

Problem: Repeat [\[link\]](#) for a negative charge.

Exercise:

Problem:

What is the maximum magnitude of the force on an aluminum rod with a $0.100\text{-}\mu\text{C}$ charge that you pass between the poles of a 1.50-T permanent magnet at a speed of 5.00 m/s ? In what direction is the force?

Solution:

$7.50 \times 10^{-7}\text{ N}$ perpendicular to both the magnetic field lines and the velocity

Exercise:

Problem:

(a) Aircraft sometimes acquire small static charges. Suppose a supersonic jet has a $0.500\text{-}\mu\text{C}$ charge and flies due west at a speed of 660 m/s over the Earth's magnetic south pole (near Earth's geographic north pole), where the $8.00 \times 10^{-5}\text{-T}$ magnetic field points straight down. What are the direction and the magnitude of the magnetic force on the plane? (b) Discuss whether the value obtained in part (a) implies this is a significant or negligible effect.

Exercise:**Problem:**

(a) A cosmic ray proton moving toward the Earth at $5.00 \times 10^7\text{ m/s}$ experiences a magnetic force of $1.70 \times 10^{-16}\text{ N}$. What is the strength of the magnetic field if there is a 45° angle between it and the proton's velocity? (b) Is the value obtained in part (a) consistent with the known strength of the Earth's magnetic field on its surface? Discuss.

Solution:

(a) $3.01 \times 10^{-5}\text{ T}$

(b) This is slightly less than the magnetic field strength of $5 \times 10^{-5}\text{ T}$ at the surface of the Earth, so it is consistent.

Exercise:**Problem:**

An electron moving at $4.00 \times 10^3\text{ m/s}$ in a 1.25-T magnetic field experiences a magnetic force of $1.40 \times 10^{-16}\text{ N}$. What angle does the velocity of the electron make with the magnetic field? There are two answers.

Exercise:

Problem:

(a) A physicist performing a sensitive measurement wants to limit the magnetic force on a moving charge in her equipment to less than 1.00×10^{-12} N. What is the greatest the charge can be if it moves at a maximum speed of 30.0 m/s in the Earth's field? (b) Discuss whether it would be difficult to limit the charge to less than the value found in (a) by comparing it with typical static electricity and noting that static is often absent.

Solution:

(a) 6.67×10^{-10} C (taking the Earth's field to be 5.00×10^{-5} T)

(b) Less than typical static, therefore difficult

Glossary

right hand rule 1 (RHR-1)

the rule to determine the direction of the magnetic force on a positive moving charge: when the thumb of the right hand points in the direction of the charge's velocity \mathbf{v} and the fingers point in the direction of the magnetic field \mathbf{B} , then the force on the charge is perpendicular and away from the palm; the force on a negative charge is perpendicular and into the palm

Lorentz force

the force on a charge moving in a magnetic field

tesla

T, the SI unit of the magnetic field strength; $1 \text{ T} = \frac{1 \text{ N}}{\text{A} \cdot \text{m}}$

magnetic force

the force on a charge produced by its motion through a magnetic field;
the Lorentz force

gauss

G, the unit of the magnetic field strength; $1 \text{ G} = 10^{-4} \text{ T}$

Force on a Moving Charge in a Magnetic Field: Examples and Applications

- Describe the effects of a magnetic field on a moving charge.
- Calculate the radius of curvature of the path of a charge that is moving in a magnetic field.

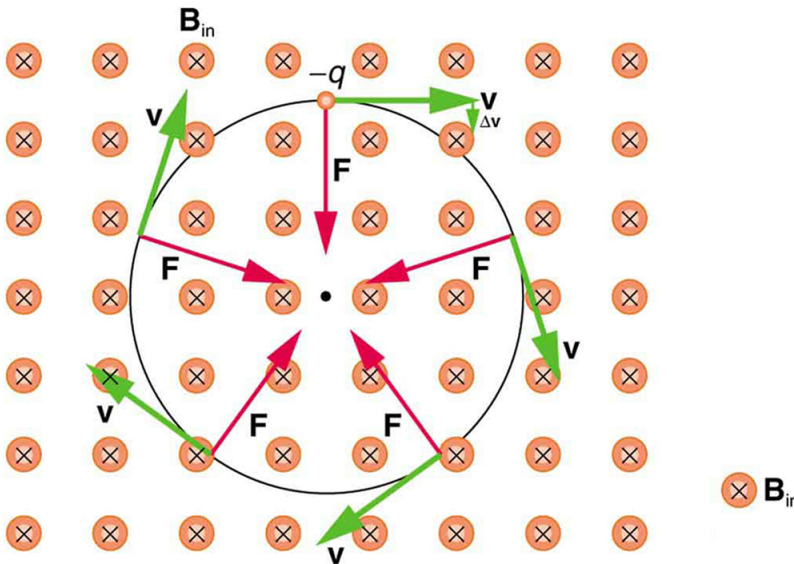
Magnetic force can cause a charged particle to move in a circular or spiral path. Cosmic rays are energetic charged particles in outer space, some of which approach the Earth. They can be forced into spiral paths by the Earth's magnetic field. Protons in giant accelerators are kept in a circular path by magnetic force. The bubble chamber photograph in [\[link\]](#) shows charged particles moving in such curved paths. The curved paths of charged particles in magnetic fields are the basis of a number of phenomena and can even be used analytically, such as in a mass spectrometer.



Trails of bubbles are produced by high-energy charged particles moving through the superheated liquid hydrogen in this artist's rendition of a bubble chamber. There is a strong magnetic field perpendicular to the page that causes the curved paths of the particles. The radius of the path can be

used to find the mass,
charge, and energy of the
particle.

So does the magnetic force cause circular motion? Magnetic force is always perpendicular to velocity, so that it does no work on the charged particle. The particle's kinetic energy and speed thus remain constant. The direction of motion is affected, but not the speed. This is typical of uniform circular motion. The simplest case occurs when a charged particle moves perpendicular to a uniform B -field, such as shown in [\[link\]](#). (If this takes place in a vacuum, the magnetic field is the dominant factor determining the motion.) Here, the magnetic force supplies the centripetal force $F_c = mv^2/r$. Noting that $\sin \theta = 1$, we see that $F = qvB$.



A negatively charged particle moves in the plane of the page in a region where the magnetic field is perpendicular into the page (represented by the small circles with x's—like the tails of arrows). The magnetic force is perpendicular to the velocity, and so velocity changes in direction but not magnitude. Uniform circular motion results.

Because the magnetic force F supplies the centripetal force F_c , we have
Equation:

$$qvB = \frac{mv^2}{r}.$$

Solving for r yields
Equation:

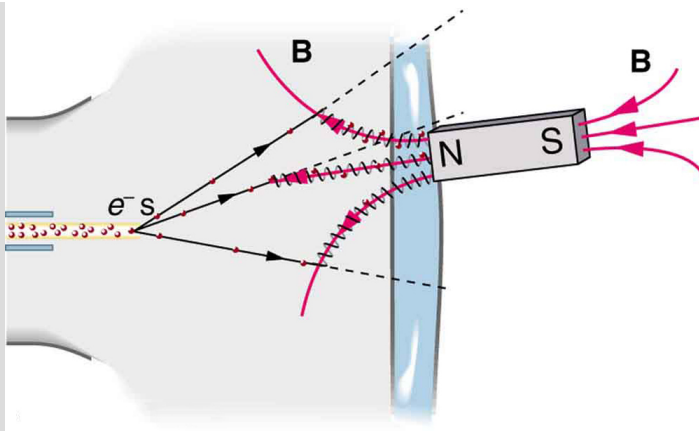
$$r = \frac{mv}{qB}.$$

Here, r is the radius of curvature of the path of a charged particle with mass m and charge q , moving at a speed v perpendicular to a magnetic field of strength B . If the velocity is not perpendicular to the magnetic field, then v is the component of the velocity perpendicular to the field. The component of the velocity parallel to the field is unaffected, since the magnetic force is zero for motion parallel to the field. This produces a spiral motion rather than a circular one.

Example:

Calculating the Curvature of the Path of an Electron Moving in a Magnetic Field: A Magnet on a TV Screen

A magnet brought near an old-fashioned TV screen such as in [\[link\]](#) (TV sets with cathode ray tubes instead of LCD screens) severely distorts its picture by altering the path of the electrons that make its phosphors glow. ***(Don't try this at home, as it will permanently magnetize and ruin the TV.)*** To illustrate this, calculate the radius of curvature of the path of an electron having a velocity of 6.00×10^7 m/s (corresponding to the accelerating voltage of about 10.0 kV used in some TVs) perpendicular to a magnetic field of strength $B = 0.500$ T (obtainable with permanent magnets).



Side view showing what happens when a magnet comes in contact with a computer monitor or TV screen. Electrons moving toward the screen spiral about magnetic field lines, maintaining the component of their velocity parallel to the field lines. This distorts the image on the screen.

Strategy

We can find the radius of curvature r directly from the equation $r = \frac{mv}{qB}$, since all other quantities in it are given or known.

Solution

Using known values for the mass and charge of an electron, along with the given values of v and B gives us

Equation:

$$\begin{aligned} r = \frac{mv}{qB} &= \frac{(9.11 \times 10^{-31} \text{ kg})(6.00 \times 10^7 \text{ m/s})}{(1.60 \times 10^{-19} \text{ C})(0.500 \text{ T})} \\ &= 6.83 \times 10^{-4} \text{ m} \end{aligned}$$

or

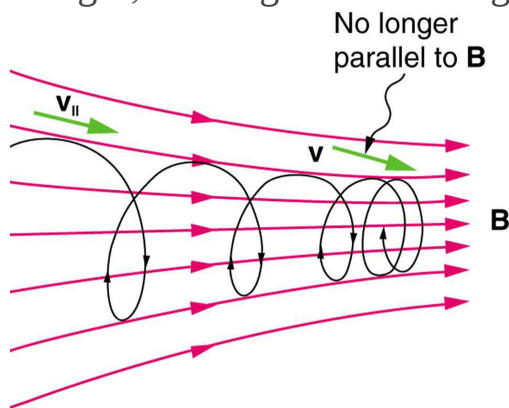
Equation:

$$r = 0.683 \text{ mm.}$$

Discussion

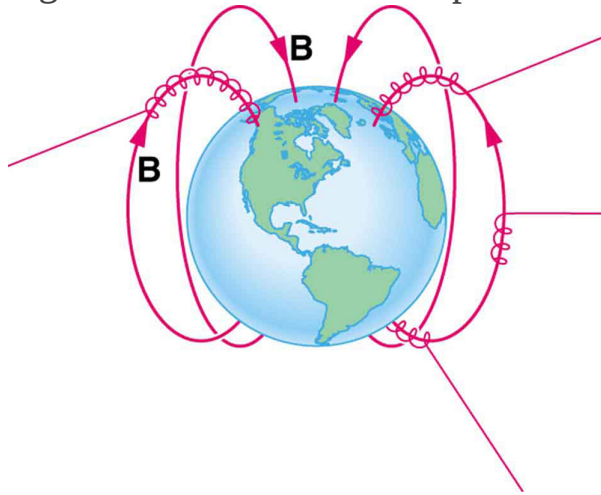
The small radius indicates a large effect. The electrons in the TV picture tube are made to move in very tight circles, greatly altering their paths and distorting the image.

[\[link\]](#) shows how electrons not moving perpendicular to magnetic field lines follow the field lines. The component of velocity parallel to the lines is unaffected, and so the charges spiral along the field lines. If field strength increases in the direction of motion, the field will exert a force to slow the charges, forming a kind of magnetic mirror, as shown below.



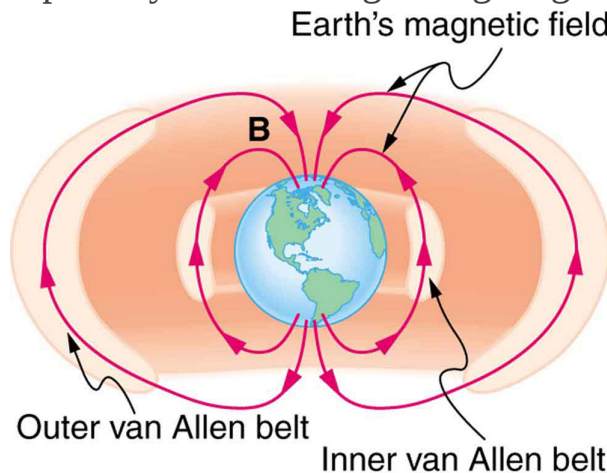
When a charged particle moves along a magnetic field line into a region where the field becomes stronger, the particle experiences a force that reduces the component of velocity parallel to the field. This force slows the motion along the field line and here reverses it, forming a “magnetic mirror.”

The properties of charged particles in magnetic fields are related to such different things as the Aurora Australis or Aurora Borealis and particle accelerators. *Charged particles approaching magnetic field lines may get trapped in spiral orbits about the lines rather than crossing them*, as seen above. Some cosmic rays, for example, follow the Earth's magnetic field lines, entering the atmosphere near the magnetic poles and causing the southern or northern lights through their ionization of molecules in the atmosphere. This glow of energized atoms and molecules is seen in [\[link\]](#). Those particles that approach middle latitudes must cross magnetic field lines, and many are prevented from penetrating the atmosphere. Cosmic rays are a component of background radiation; consequently, they give a higher radiation dose at the poles than at the equator.



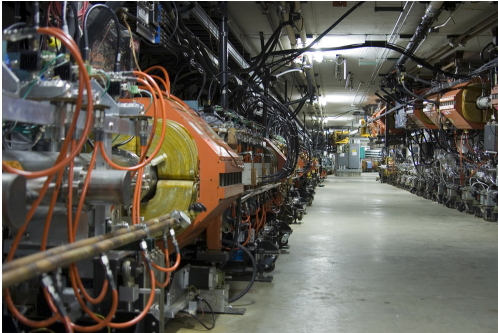
Energetic electrons and protons, components of cosmic rays, from the Sun and deep outer space often follow the Earth's magnetic field lines rather than cross them. (Recall that the Earth's north magnetic pole is really a south pole in terms of a bar magnet.)

Some incoming charged particles become trapped in the Earth's magnetic field, forming two belts above the atmosphere known as the Van Allen radiation belts after the discoverer James A. Van Allen, an American astrophysicist. (See [\[link\]](#).) Particles trapped in these belts form radiation fields (similar to nuclear radiation) so intense that manned space flights avoid them and satellites with sensitive electronics are kept out of them. In the few minutes it took lunar missions to cross the Van Allen radiation belts, astronauts received radiation doses more than twice the allowed annual exposure for radiation workers. Other planets have similar belts, especially those having strong magnetic fields like Jupiter.



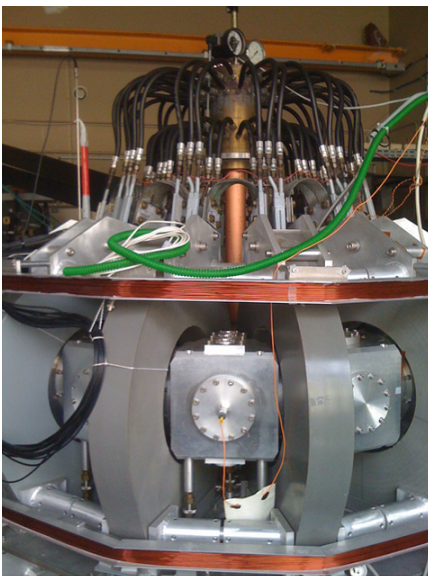
The Van Allen radiation belts are two regions in which energetic charged particles are trapped in the Earth's magnetic field. One belt lies about 300 km above the Earth's surface, the other about 16,000 km. Charged particles in these belts migrate along magnetic field lines and are partially reflected away from the poles by the stronger fields there. The charged particles that enter the atmosphere are replenished by the Sun and sources in deep outer space.

Back on Earth, we have devices that employ magnetic fields to contain charged particles. Among them are the giant particle accelerators that have been used to explore the substructure of matter. (See [\[link\]](#).) Magnetic fields not only control the direction of the charged particles, they also are used to focus particles into beams and overcome the repulsion of like charges in these beams.

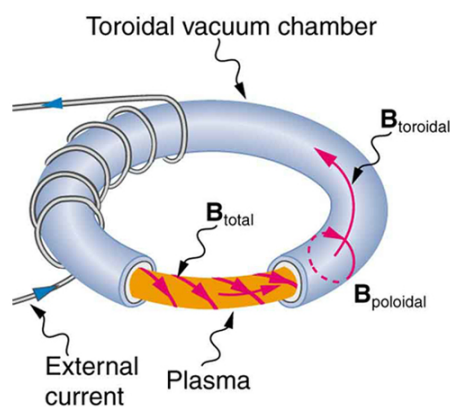


The Fermilab facility in Illinois has a large particle accelerator (the most powerful in the world until 2008) that employs magnetic fields (magnets seen here in orange) to contain and direct its beam. This and other accelerators have been in use for several decades and have allowed us to discover some of the laws underlying all matter. (credit: ammcgrim, Flickr)

Thermonuclear fusion (like that occurring in the Sun) is a hope for a future clean energy source. One of the most promising devices is the *tokamak*, which uses magnetic fields to contain (or trap) and direct the reactive charged particles. (See [\[link\]](#).) Less exotic, but more immediately practical, amplifiers in microwave ovens use a magnetic field to contain oscillating electrons. These oscillating electrons generate the microwaves sent into the oven.



(a)



(b)

Tokamaks such as the one shown in the figure are being studied with the goal of economical production of energy by nuclear fusion. Magnetic fields in the doughnut-shaped device contain and direct the reactive charged particles. (credit: David Mellis, Flickr)

Mass spectrometers have a variety of designs, and many use magnetic fields to measure mass. The curvature of a charged particle's path in the field is related to its mass and is measured to obtain mass information. (See [More Applications of Magnetism](#).) Historically, such techniques were employed in the first direct observations of electron charge and mass. Today, mass

spectrometers (sometimes coupled with gas chromatographs) are used to determine the make-up and sequencing of large biological molecules.

Section Summary

- Magnetic force can supply centripetal force and cause a charged particle to move in a circular path of radius

Equation:

$$r = \frac{mv}{qB},$$

where v is the component of the velocity perpendicular to B for a charged particle with mass m and charge q .

Conceptual Questions

Exercise:

Problem:

How can the motion of a charged particle be used to distinguish between a magnetic and an electric field?

Exercise:

Problem:

High-velocity charged particles can damage biological cells and are a component of radiation exposure in a variety of locations ranging from research facilities to natural background. Describe how you could use a magnetic field to shield yourself.

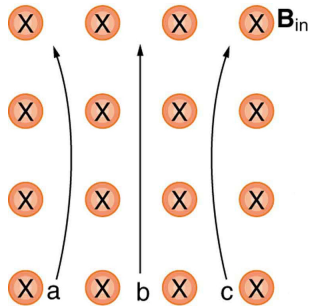
Exercise:

Problem:

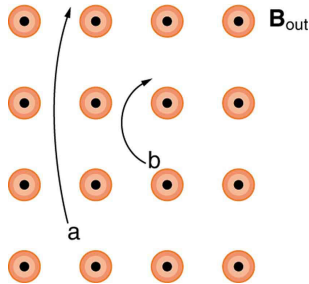
If a cosmic ray proton approaches the Earth from outer space along a line toward the center of the Earth that lies in the plane of the equator, in what direction will it be deflected by the Earth's magnetic field? What about an electron? A neutron?

Exercise:

Problem: What are the signs of the charges on the particles in [\[link\]](#)?

**Exercise:****Problem:**

Which of the particles in [\[link\]](#) has the greatest velocity, assuming they have identical charges and masses?

**Exercise:****Problem:**

Which of the particles in [\[link\]](#) has the greatest mass, assuming all have identical charges and velocities?

Exercise:

Problem:

While operating, a high-precision TV monitor is placed on its side during maintenance. The image on the monitor changes color and blurs slightly. Discuss the possible relation of these effects to the Earth's magnetic field.

Problems & Exercises

If you need additional support for these problems, see [More Applications of Magnetism](#).

Exercise:**Problem:**

A cosmic ray electron moves at 7.50×10^6 m/s perpendicular to the Earth's magnetic field at an altitude where field strength is 1.00×10^{-5} T. What is the radius of the circular path the electron follows?

Solution:

4.27 m

Exercise:**Problem:**

A proton moves at 7.50×10^7 m/s perpendicular to a magnetic field. The field causes the proton to travel in a circular path of radius 0.800 m. What is the field strength?

Exercise:

Problem:

(a) Viewers of *Star Trek* hear of an antimatter drive on the Starship *Enterprise*. One possibility for such a futuristic energy source is to store antimatter charged particles in a vacuum chamber, circulating in a magnetic field, and then extract them as needed. Antimatter annihilates with normal matter, producing pure energy. What strength magnetic field is needed to hold antiprotons, moving at 5.00×10^7 m/s in a circular path 2.00 m in radius? Antiprotons have the same mass as protons but the opposite (negative) charge. (b) Is this field strength obtainable with today's technology or is it a futuristic possibility?

Solution:

(a) 0.261 T

(b) This strength is definitely obtainable with today's technology. Magnetic field strengths of 0.500 T are obtainable with permanent magnets.

Exercise:**Problem:**

(a) An oxygen-16 ion with a mass of 2.66×10^{-26} kg travels at 5.00×10^6 m/s perpendicular to a 1.20-T magnetic field, which makes it move in a circular arc with a 0.231-m radius. What positive charge is on the ion? (b) What is the ratio of this charge to the charge of an electron? (c) Discuss why the ratio found in (b) should be an integer.

Exercise:**Problem:**

What radius circular path does an electron travel if it moves at the same speed and in the same magnetic field as the proton in [\[link\]](#)?

Solution:

$$4.36 \times 10^{-4} \text{ m}$$

Exercise:**Problem:**

A velocity selector in a mass spectrometer uses a 0.100-T magnetic field. (a) What electric field strength is needed to select a speed of $4.00 \times 10^6 \text{ m/s}$? (b) What is the voltage between the plates if they are separated by 1.00 cm?

Exercise:**Problem:**

An electron in a TV CRT moves with a speed of $6.00 \times 10^7 \text{ m/s}$, in a direction perpendicular to the Earth's field, which has a strength of $5.00 \times 10^{-5} \text{ T}$. (a) What strength electric field must be applied perpendicular to the Earth's field to make the electron moves in a straight line? (b) If this is done between plates separated by 1.00 cm, what is the voltage applied? (Note that TVs are usually surrounded by a ferromagnetic material to shield against external magnetic fields and avoid the need for such a correction.)

Solution:

(a) 3.00 kV/m

(b) 30.0 V

Exercise:

Problem:

(a) At what speed will a proton move in a circular path of the same radius as the electron in [\[link\]](#)? (b) What would the radius of the path be if the proton had the same speed as the electron? (c) What would the radius be if the proton had the same kinetic energy as the electron? (d) The same momentum?

Exercise:**Problem:**

A mass spectrometer is being used to separate common oxygen-16 from the much rarer oxygen-18, taken from a sample of old glacial ice. (The relative abundance of these oxygen isotopes is related to climatic temperature at the time the ice was deposited.) The ratio of the masses of these two ions is 16 to 18, the mass of oxygen-16 is 2.66×10^{-26} kg, and they are singly charged and travel at 5.00×10^6 m/s in a 1.20-T magnetic field. What is the separation between their paths when they hit a target after traversing a semicircle?

Solution:

0.173 m

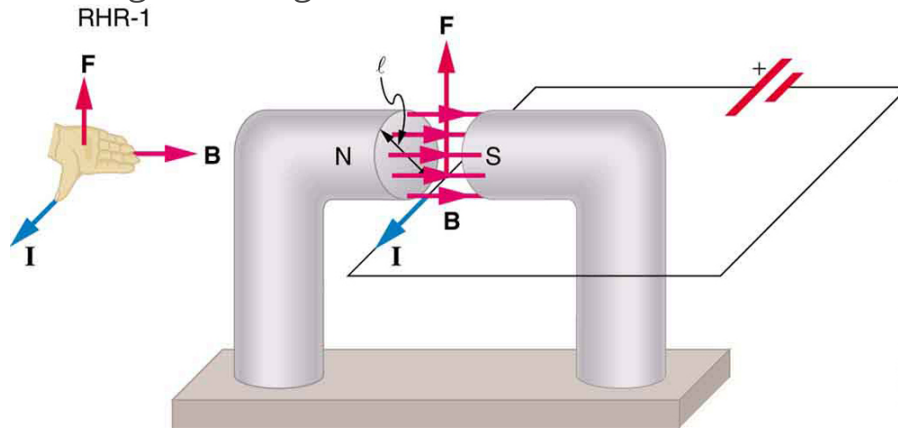
Exercise:**Problem:**

(a) Triply charged uranium-235 and uranium-238 ions are being separated in a mass spectrometer. (The much rarer uranium-235 is used as reactor fuel.) The masses of the ions are 3.90×10^{-25} kg and 3.95×10^{-25} kg, respectively, and they travel at 3.00×10^5 m/s in a 0.250-T field. What is the separation between their paths when they hit a target after traversing a semicircle? (b) Discuss whether this distance between their paths seems to be big enough to be practical in the separation of uranium-235 from uranium-238.

Magnetic Force on a Current-Carrying Conductor

- Describe the effects of a magnetic force on a current-carrying conductor.
- Calculate the magnetic force on a current-carrying conductor.

Because charges ordinarily cannot escape a conductor, the magnetic force on charges moving in a conductor is transmitted to the conductor itself.



The magnetic field exerts a force on a current-carrying wire in a direction given by the right hand rule 1 (the same direction as that on the individual moving charges). This force can easily be large enough to move the wire, since typical currents consist of very large numbers of moving charges.

We can derive an expression for the magnetic force on a current by taking a sum of the magnetic forces on individual charges. (The forces add because they are in the same direction.) The force on an individual charge moving at the drift velocity v_d is given by $F = qv_d B \sin \theta$. Taking B to be uniform over a length of wire l and zero elsewhere, the total magnetic force on the wire is then $F = (qv_d B \sin \theta)(N)$, where N is the number of charge carriers in the section of wire of length l . Now, $N = nV$, where n is the number of charge carriers per unit volume and V is the volume of wire in the field. Noting that $V = Al$, where A is the cross-sectional area of the

wire, then the force on the wire is $F = (qv_d B \sin \theta)(nAl)$. Gathering terms,

Equation:

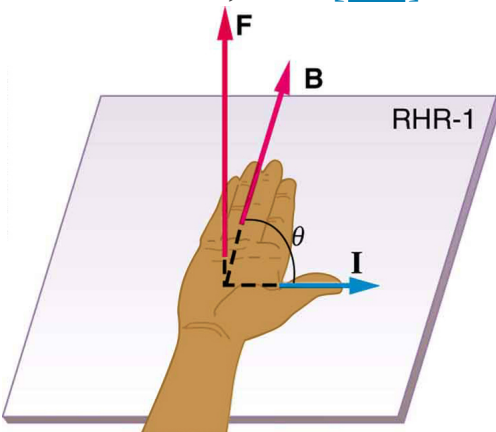
$$F = (nqAv_d)lB \sin \theta.$$

Because $nqAv_d = I$ (see [Current](#)),

Equation:

$$F = IlB \sin \theta$$

is the equation for *magnetic force on a length l of wire carrying a current I in a uniform magnetic field B* , as shown in [\[link\]](#). If we divide both sides of this expression by l , we find that the magnetic force per unit length of wire in a uniform field is $\frac{F}{l} = IB \sin \theta$. The direction of this force is given by RHR-1, with the thumb in the direction of the current I . Then, with the fingers in the direction of B , a perpendicular to the palm points in the direction of F , as in [\[link\]](#).



$$F = IlB \sin \theta$$

$\mathbf{F} \perp$ plane of \mathbf{I} and \mathbf{B}

The force on a current-carrying wire in a magnetic field is $F = IlB \sin \theta$. Its

direction is given by
RHR-1.

Example:

Calculating Magnetic Force on a Current-Carrying Wire: A Strong Magnetic Field

Calculate the force on the wire shown in [\[link\]](#), given $B = 1.50 \text{ T}$, $l = 5.00 \text{ cm}$, and $I = 20.0 \text{ A}$.

Strategy

The force can be found with the given information by using $F = IlB \sin \theta$ and noting that the angle θ between I and B is 90° , so that $\sin \theta = 1$.

Solution

Entering the given values into $F = IlB \sin \theta$ yields

Equation:

$$F = IlB \sin \theta = (20.0 \text{ A})(0.0500 \text{ m})(1.50 \text{ T})(1).$$

The units for tesla are $1 \text{ T} = \frac{\text{N}}{\text{A}\cdot\text{m}}$; thus,

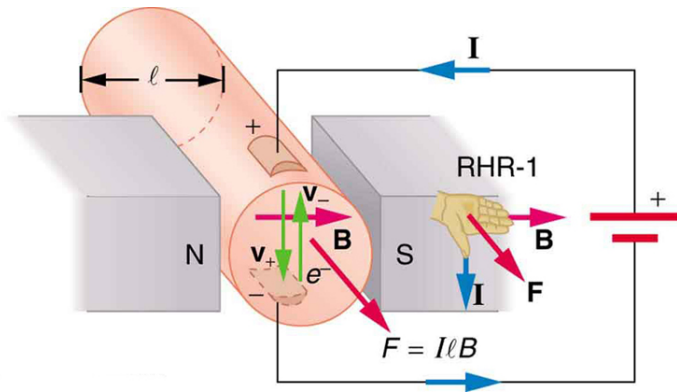
Equation:

$$F = 1.50 \text{ N}.$$

Discussion

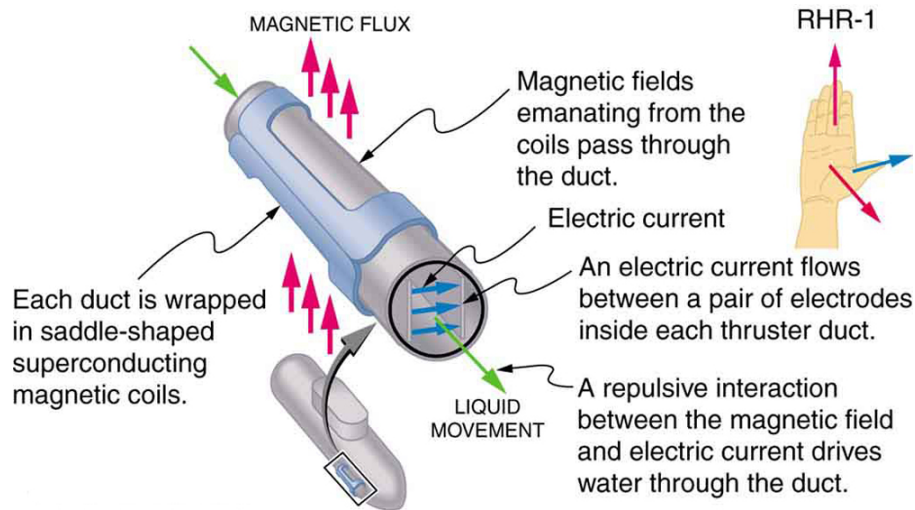
This large magnetic field creates a significant force on a small length of wire.

Magnetic force on current-carrying conductors is used to convert electric energy to work. (Motors are a prime example—they employ loops of wire and are considered in the next section.) Magnetohydrodynamics (MHD) is the technical name given to a clever application where magnetic force pumps fluids without moving mechanical parts. (See [\[link\]](#).)



Magnetohydrodynamics. The magnetic force on the current passed through this fluid can be used as a nonmechanical pump.

A strong magnetic field is applied across a tube and a current is passed through the fluid at right angles to the field, resulting in a force on the fluid parallel to the tube axis as shown. The absence of moving parts makes this attractive for moving a hot, chemically active substance, such as the liquid sodium employed in some nuclear reactors. Experimental artificial hearts are testing with this technique for pumping blood, perhaps circumventing the adverse effects of mechanical pumps. (Cell membranes, however, are affected by the large fields needed in MHD, delaying its practical application in humans.) MHD propulsion for nuclear submarines has been proposed, because it could be considerably quieter than conventional propeller drives. The deterrent value of nuclear submarines is based on their ability to hide and survive a first or second nuclear strike. As we slowly disassemble our nuclear weapons arsenals, the submarine branch will be the last to be decommissioned because of this ability (See [\[link\]](#).) Existing MHD drives are heavy and inefficient—much development work is needed.



An MHD propulsion system in a nuclear submarine could produce significantly less turbulence than propellers and allow it to run more silently. The development of a silent drive submarine was dramatized in the book and the film *The Hunt for Red October*.

Section Summary

- The magnetic force on current-carrying conductors is given by **Equation:**

$$F = IlB \sin \theta,$$

where I is the current, l is the length of a straight conductor in a uniform magnetic field B , and θ is the angle between I and B . The force follows RHR-1 with the thumb in the direction of I .

Conceptual Questions

Exercise:

Problem:

Draw a sketch of the situation in [\[link\]](#) showing the direction of electrons carrying the current, and use RHR-1 to verify the direction of the force on the wire.

Exercise:**Problem:**

Verify that the direction of the force in an MHD drive, such as that in [\[link\]](#), does not depend on the sign of the charges carrying the current across the fluid.

Exercise:**Problem:**

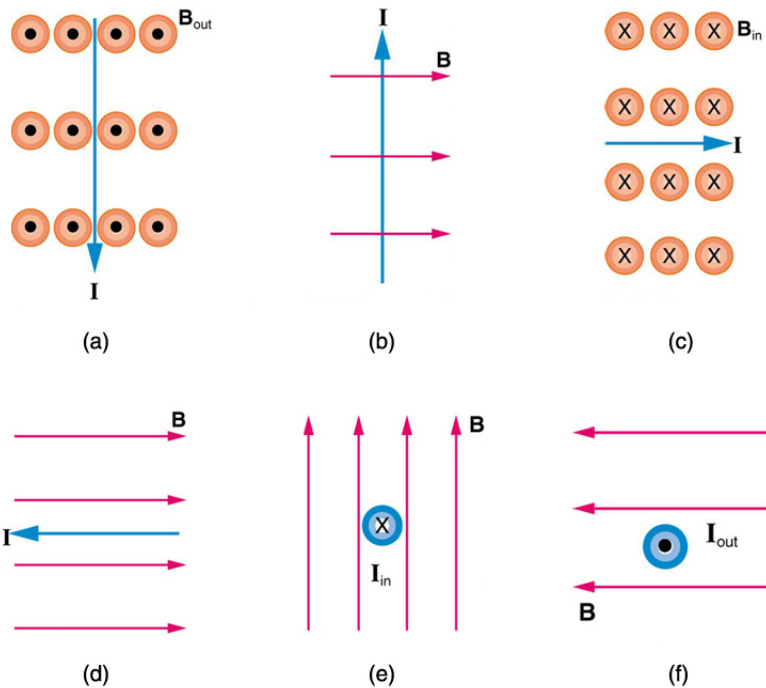
Why would a magnetohydrodynamic drive work better in ocean water than in fresh water? Also, why would superconducting magnets be desirable?

Exercise:**Problem:**

Which is more likely to interfere with compass readings, AC current in your refrigerator or DC current when you start your car? Explain.

Problems & Exercises**Exercise:****Problem:**

What is the direction of the magnetic force on the current in each of the six cases in [\[link\]](#)?



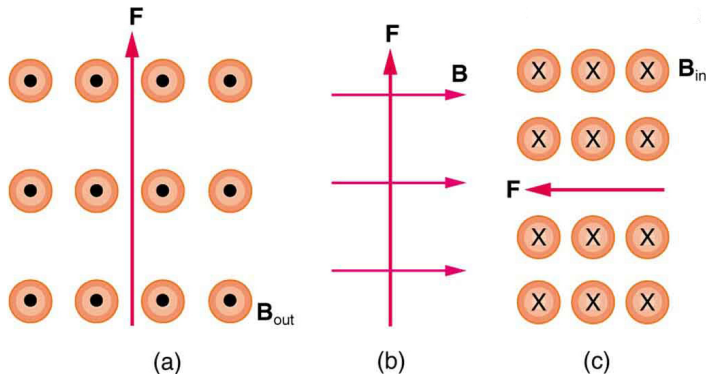
Solution:

- (a) west (left)
- (b) into page
- (c) north (up)
- (d) no force
- (e) east (right)
- (f) south (down)

Exercise:

Problem:

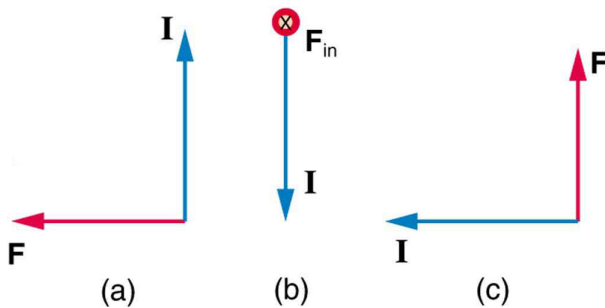
What is the direction of a current that experiences the magnetic force shown in each of the three cases in [\[link\]](#), assuming the current runs perpendicular to B ?



Exercise:

Problem:

What is the direction of the magnetic field that produces the magnetic force shown on the currents in each of the three cases in [\[link\]](#), assuming \mathbf{B} is perpendicular to \mathbf{I} ?



Solution:

(a) into page

(b) west (left)

(c) out of page

Exercise:

Problem:

(a) What is the force per meter on a lightning bolt at the equator that carries 20,000 A perpendicular to the Earth's 3.00×10^{-5} -T field? (b) What is the direction of the force if the current is straight up and the Earth's field direction is due north, parallel to the ground?

Exercise:**Problem:**

(a) A DC power line for a light-rail system carries 1000 A at an angle of 30.0° to the Earth's 5.00×10^{-5} -T field. What is the force on a 100-m section of this line? (b) Discuss practical concerns this presents, if any.

Solution:

(a) 2.50 N

(b) This is about half a pound of force per 100 m of wire, which is much less than the weight of the wire itself. Therefore, it does not cause any special concerns.

Exercise:**Problem:**

What force is exerted on the water in an MHD drive utilizing a 25.0-cm-diameter tube, if 100-A current is passed across the tube that is perpendicular to a 2.00-T magnetic field? (The relatively small size of this force indicates the need for very large currents and magnetic fields to make practical MHD drives.)

Exercise:**Problem:**

A wire carrying a 30.0-A current passes between the poles of a strong magnet that is perpendicular to its field and experiences a 2.16-N force on the 4.00 cm of wire in the field. What is the average field strength?

Solution:

1.80 T

Exercise:

Problem:

(a) A 0.750-m-long section of cable carrying current to a car starter motor makes an angle of 60° with the Earth's 5.50×10^{-5} T field. What is the current when the wire experiences a force of 7.00×10^{-3} N? (b) If you run the wire between the poles of a strong horseshoe magnet, subjecting 5.00 cm of it to a 1.75-T field, what force is exerted on this segment of wire?

Exercise:**Problem:**

(a) What is the angle between a wire carrying an 8.00-A current and the 1.20-T field it is in if 50.0 cm of the wire experiences a magnetic force of 2.40 N? (b) What is the force on the wire if it is rotated to make an angle of 90° with the field?

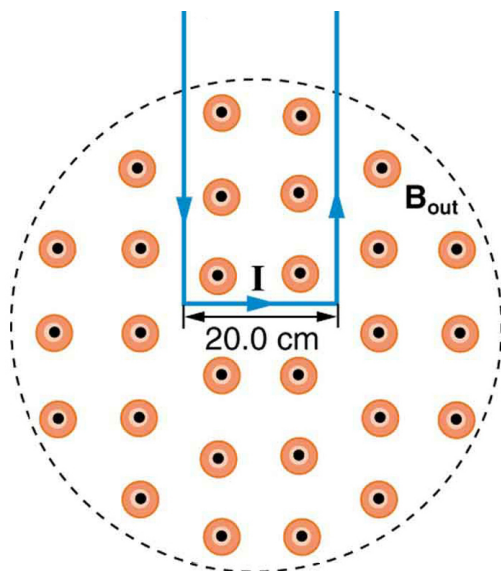
Solution:

(a) 30°

(b) 4.80 N

Exercise:**Problem:**

The force on the rectangular loop of wire in the magnetic field in [\[link\]](#) can be used to measure field strength. The field is uniform, and the plane of the loop is perpendicular to the field. (a) What is the direction of the magnetic force on the loop? Justify the claim that the forces on the sides of the loop are equal and opposite, independent of how much of the loop is in the field and do not affect the net force on the loop. (b) If a current of 5.00 A is used, what is the force per tesla on the 20.0-cm-wide loop?

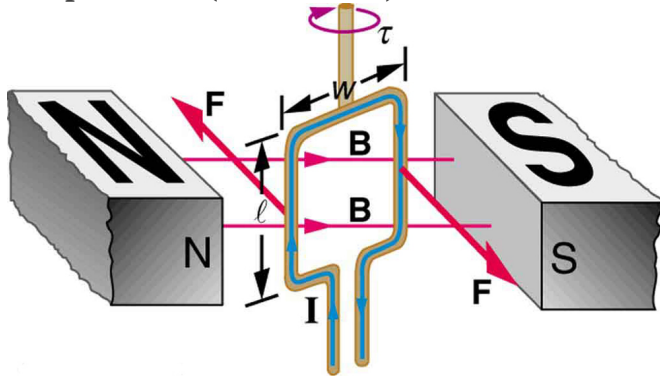


A rectangular loop of wire carrying a current is perpendicular to a magnetic field. The field is uniform in the region shown and is zero outside that region.

Torque on a Current Loop: Motors and Meters

- Describe how motors and meters work in terms of torque on a current loop.
- Calculate the torque on a current-carrying loop in a magnetic field.

Motors are the most common application of magnetic force on current-carrying wires. Motors have loops of wire in a magnetic field. When current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft. Electrical energy is converted to mechanical work in the process. (See [\[link\]](#).)



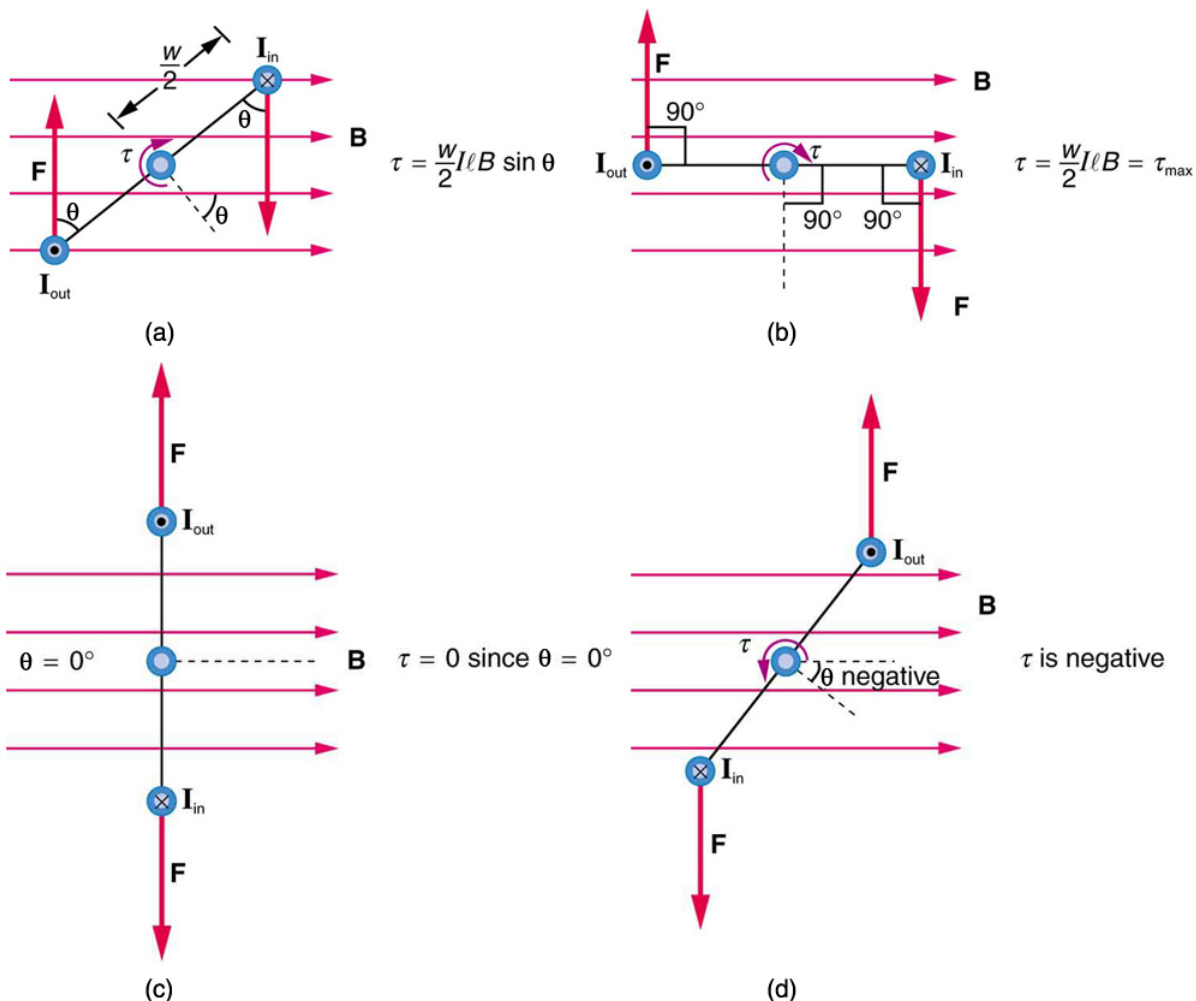
Torque on a current loop. A current-carrying loop of wire attached to a vertically rotating shaft feels magnetic forces that produce a clockwise torque as viewed from above.

Let us examine the force on each segment of the loop in [\[link\]](#) to find the torques produced about the axis of the vertical shaft. (This will lead to a useful equation for the torque on the loop.) We take the magnetic field to be uniform over the rectangular loop, which has width w and height l . First, we note that the forces on the top and bottom segments are vertical and, therefore, parallel to the shaft, producing no torque. Those vertical forces are equal in magnitude and opposite in direction, so that they also produce no net force on the loop. [\[link\]](#) shows views of the loop from above. Torque

is defined as $\tau = rF \sin \theta$, where F is the force, r is the distance from the pivot that the force is applied, and θ is the angle between r and F . As seen in [link](a), right hand rule 1 gives the forces on the sides to be equal in magnitude and opposite in direction, so that the net force is again zero. However, each force produces a clockwise torque. Since $r = w/2$, the torque on each vertical segment is $(w/2)F \sin \theta$, and the two add to give a total torque.

Equation:

$$\tau = \frac{w}{2} F \sin \theta + \frac{w}{2} F \sin \theta = wF \sin \theta$$



Top views of a current-carrying loop in a magnetic field. (a) The equation for torque is derived using this view. Note that the perpendicular to the loop makes an angle θ with the field that is the

same as the angle between $w/2$ and \mathbf{F} . (b) The maximum torque occurs when θ is a right angle and $\sin \theta = 1$. (c) Zero (minimum) torque occurs when θ is zero and $\sin \theta = 0$. (d) The torque reverses once the loop rotates past $\theta = 0$.

Now, each vertical segment has a length l that is perpendicular to B , so that the force on each is $F = IlB$. Entering F into the expression for torque yields

Equation:

$$\tau = wIlB \sin \theta.$$

If we have a multiple loop of N turns, we get N times the torque of one loop. Finally, note that the area of the loop is $A = wl$; the expression for the torque becomes

Equation:

$$\tau = NIAB \sin \theta.$$

This is the torque on a current-carrying loop in a uniform magnetic field. This equation can be shown to be valid for a loop of any shape. The loop carries a current I , has N turns, each of area A , and the perpendicular to the loop makes an angle θ with the field B . The net force on the loop is zero.

Example:

Calculating Torque on a Current-Carrying Loop in a Strong Magnetic Field

Find the maximum torque on a 100-turn square loop of a wire of 10.0 cm on a side that carries 15.0 A of current in a 2.00-T field.

Strategy

Torque on the loop can be found using $\tau = NIAB \sin \theta$. Maximum torque occurs when $\theta = 90^\circ$ and $\sin \theta = 1$.

Solution

For $\sin \theta = 1$, the maximum torque is

Equation:

$$\tau_{\max} = NIAB.$$

Entering known values yields

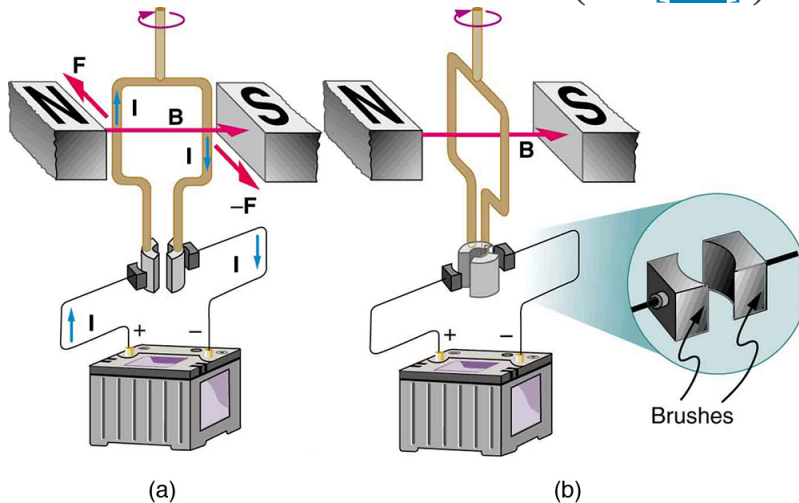
Equation:

$$\begin{aligned}\tau_{\max} &= (100)(15.0 \text{ A})(0.100 \text{ m}^2)(2.00 \text{ T}) \\ &= 30.0 \text{ N} \cdot \text{m}.\end{aligned}$$

Discussion

This torque is large enough to be useful in a motor.

The torque found in the preceding example is the maximum. As the coil rotates, the torque decreases to zero at $\theta = 0$. The torque then *reverses* its direction once the coil rotates past $\theta = 0$. (See [\[link\]](#)(d).) This means that, unless we do something, the coil will oscillate back and forth about equilibrium at $\theta = 0$. To get the coil to continue rotating in the same direction, we can reverse the current as it passes through $\theta = 0$ with automatic switches called *brushes*. (See [\[link\]](#).)

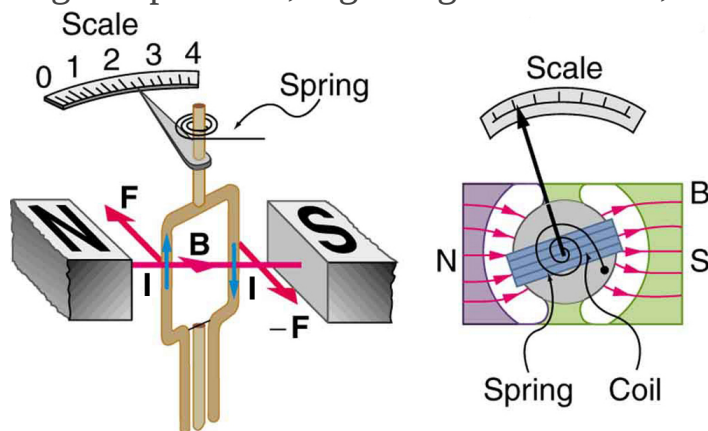


(a) As the angular momentum of the coil carries it through $\theta = 0$, the brushes reverse

the current to keep the torque clockwise. (b)

The coil will rotate continuously in the clockwise direction, with the current reversing each half revolution to maintain the clockwise torque.

Meters, such as those in analog fuel gauges on a car, are another common application of magnetic torque on a current-carrying loop. [\[link\]](#) shows that a meter is very similar in construction to a motor. The meter in the figure has its magnets shaped to limit the effect of θ by making B perpendicular to the loop over a large angular range. Thus the torque is proportional to I and not θ . A linear spring exerts a counter-torque that balances the current-produced torque. This makes the needle deflection proportional to I . If an exact proportionality cannot be achieved, the gauge reading can be calibrated. To produce a galvanometer for use in analog voltmeters and ammeters that have a low resistance and respond to small currents, we use a large loop area A , high magnetic field B , and low-resistance coils.



Meters are very similar to motors but only rotate through a part of a revolution. The magnetic poles of this meter are shaped to keep the component of B perpendicular to the loop constant, so that the torque does not depend on θ and the deflection

against the return spring is proportional only to the current I .

Section Summary

- The torque τ on a current-carrying loop of any shape in a uniform magnetic field. is

Equation:

$$\tau = NIAB \sin \theta,$$

where N is the number of turns, I is the current, A is the area of the loop, B is the magnetic field strength, and θ is the angle between the perpendicular to the loop and the magnetic field.

Conceptual Questions

Exercise:

Problem:

Draw a diagram and use RHR-1 to show that the forces on the top and bottom segments of the motor's current loop in [\[link\]](#) are vertical and produce no torque about the axis of rotation.

Problems & Exercises

Exercise:

Problem:

(a) By how many percent is the torque of a motor decreased if its permanent magnets lose 5.0% of their strength? (b) How many percent would the current need to be increased to return the torque to original values?

Solution:

(a) τ decreases by 5.00% if B decreases by 5.00%

(b) 5.26% increase

Exercise:

Problem:

(a) What is the maximum torque on a 150-turn square loop of wire 18.0 cm on a side that carries a 50.0-A current in a 1.60-T field? (b) What is the torque when θ is 10.9° ?

Exercise:

Problem:

Find the current through a loop needed to create a maximum torque of $9.00 \text{ N} \cdot \text{m}$. The loop has 50 square turns that are 15.0 cm on a side and is in a uniform 0.800-T magnetic field.

Solution:

10.0 A

Exercise:

Problem:

Calculate the magnetic field strength needed on a 200-turn square loop 20.0 cm on a side to create a maximum torque of $300 \text{ N} \cdot \text{m}$ if the loop is carrying 25.0 A.

Exercise:

Problem:

Since the equation for torque on a current-carrying loop is $\tau = NIAB \sin \theta$, the units of $\text{N} \cdot \text{m}$ must equal units of $\text{A} \cdot \text{m}^2 \text{ T}$. Verify this.

Solution:

$$A \cdot m^2 \cdot T = A \cdot m^2 \left(\frac{N}{A \cdot m} \right) = N \cdot m.$$

Exercise:

Problem:

(a) At what angle θ is the torque on a current loop 90.0% of maximum? (b) 50.0% of maximum? (c) 10.0% of maximum?

Exercise:

Problem:

A proton has a magnetic field due to its spin on its axis. The field is similar to that created by a circular current loop 0.650×10^{-15} m in radius with a current of 1.05×10^4 A (no kidding). Find the maximum torque on a proton in a 2.50-T field. (This is a significant torque on a small particle.)

Solution:

$$3.48 \times 10^{-26} \text{ N} \cdot \text{m}$$

Exercise:

Problem:

(a) A 200-turn circular loop of radius 50.0 cm is vertical, with its axis on an east-west line. A current of 100 A circulates clockwise in the loop when viewed from the east. The Earth's field here is due north, parallel to the ground, with a strength of 3.00×10^{-5} T. What are the direction and magnitude of the torque on the loop? (b) Does this device have any practical applications as a motor?

Exercise:

Problem:

Repeat [\[link\]](#), but with the loop lying flat on the ground with its current circulating counterclockwise (when viewed from above) in a location where the Earth's field is north, but at an angle 45.0° below the horizontal and with a strength of 6.00×10^{-5} T.

Solution:

(a) $0.666 \text{ N} \cdot \text{m}$ west

(b) This is not a very significant torque, so practical use would be limited. Also, the current would need to be alternated to make the loop rotate (otherwise it would oscillate).

Glossary

motor

loop of wire in a magnetic field; when current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft; electrical energy is converted to mechanical work in the process

meter

common application of magnetic torque on a current-carrying loop that is very similar in construction to a motor; by design, the torque is proportional to I and not θ , so the needle deflection is proportional to the current

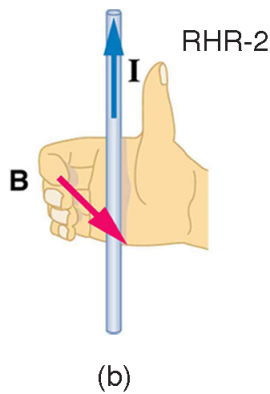
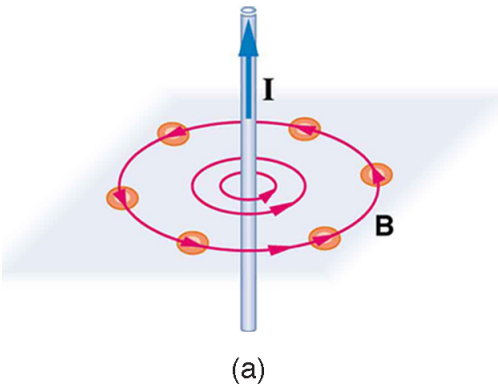
Magnetic Fields Produced by Currents: Ampere's Law

- Calculate current that produces a magnetic field.
- Use the right hand rule 2 to determine the direction of current or the direction of magnetic field loops.

How much current is needed to produce a significant magnetic field, perhaps as strong as the Earth's field? Surveyors will tell you that overhead electric power lines create magnetic fields that interfere with their compass readings. Indeed, when Oersted discovered in 1820 that a current in a wire affected a compass needle, he was not dealing with extremely large currents. How does the shape of wires carrying current affect the shape of the magnetic field created? We noted earlier that a current loop created a magnetic field similar to that of a bar magnet, but what about a straight wire or a toroid (doughnut)? How is the direction of a current-created field related to the direction of the current? Answers to these questions are explored in this section, together with a brief discussion of the law governing the fields created by currents.

Magnetic Field Created by a Long Straight Current-Carrying Wire: Right Hand Rule 2

Magnetic fields have both direction and magnitude. As noted before, one way to explore the direction of a magnetic field is with compasses, as shown for a long straight current-carrying wire in [\[link\]](#). Hall probes can determine the magnitude of the field. The field around a long straight wire is found to be in circular loops. The **right hand rule 2** (RHR-2) emerges from this exploration and is valid for any current segment—*point the thumb in the direction of the current, and the fingers curl in the direction of the magnetic field loops* created by it.



(a) Compasses placed near a long straight current-carrying wire indicate that field lines form circular loops centered on the wire. (b) Right hand rule 2 states that, if the right hand thumb points in the direction of the current, the fingers curl in the direction of the field. This rule is consistent with the field mapped for the long straight wire and is valid for any current segment.

The **magnetic field strength (magnitude) produced by a long straight current-carrying wire** is found by experiment to be
Equation:

$$B = \frac{\mu_0 I}{2\pi r} \text{ (long straight wire),}$$

where I is the current, r is the shortest distance to the wire, and the constant $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ is the **permeability of free space**. (μ_0 is one of the basic constants in nature. We will see later that μ_0 is related to the speed of light.) Since the wire is very long, the magnitude of the field depends only on distance from the wire r , not on position along the wire.

Example:

Calculating Current that Produces a Magnetic Field

Find the current in a long straight wire that would produce a magnetic field twice the strength of the Earth's at a distance of 5.0 cm from the wire.

Strategy

The Earth's field is about $5.0 \times 10^{-5} \text{ T}$, and so here B due to the wire is taken to be $1.0 \times 10^{-4} \text{ T}$. The equation $B = \frac{\mu_0 I}{2\pi r}$ can be used to find I , since all other quantities are known.

Solution

Solving for I and entering known values gives

Equation:

$$\begin{aligned} I &= \frac{2\pi r B}{\mu_0} = \frac{2\pi(5.0 \times 10^{-2} \text{ m})(1.0 \times 10^{-4} \text{ T})}{4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}} \\ &= 25 \text{ A.} \end{aligned}$$

Discussion

So a moderately large current produces a significant magnetic field at a distance of 5.0 cm from a long straight wire. Note that the answer is stated to only two digits, since the Earth's field is specified to only two digits in this example.

Ampere's Law and Others

The magnetic field of a long straight wire has more implications than you might at first suspect. *Each segment of current produces a magnetic field like that of a long straight wire, and the total field of any shape current is the vector sum of the fields due to each segment.* The formal statement of the direction and magnitude of the field due to each segment is called the **Biot-Savart law**. Integral calculus is needed to sum the field for an arbitrary shape current. This results in a more complete law, called **Ampere's law**, which relates magnetic field and current in a general way. Ampere's law in turn is a part of **Maxwell's equations**, which give a complete theory of all electromagnetic phenomena. Considerations of how Maxwell's equations appear to different observers led to the modern theory of relativity, and the realization that electric and magnetic fields are different manifestations of the same thing. Most of this is beyond the scope of this text in both mathematical level, requiring calculus, and in the amount of space that can be devoted to it. But for the interested student, and particularly for those who continue in physics, engineering, or similar pursuits, delving into these matters further will reveal descriptions of nature that are elegant as well as profound. In this text, we shall keep the general features in mind, such as RHR-2 and the rules for magnetic field lines listed in [Magnetic Fields and Magnetic Field Lines](#), while concentrating on the fields created in certain important situations.

Note:

Making Connections: Relativity

Hearing all we do about Einstein, we sometimes get the impression that he invented relativity out of nothing. On the contrary, one of Einstein's motivations was to solve difficulties in knowing how different observers see magnetic and electric fields.

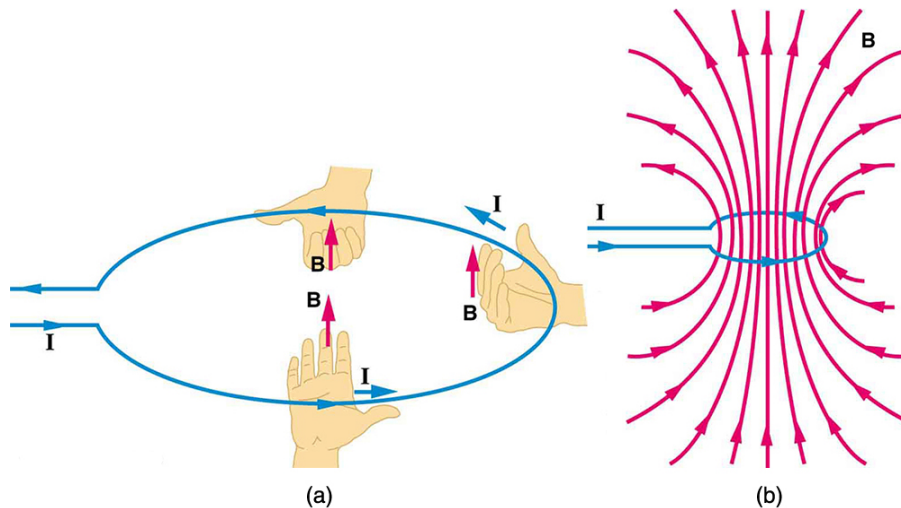
Magnetic Field Produced by a Current-Carrying Circular Loop

The magnetic field near a current-carrying loop of wire is shown in [\[link\]](#). Both the direction and the magnitude of the magnetic field produced by a current-carrying loop are complex. RHR-2 can be used to give the direction of the field near the loop, but mapping with compasses and the rules about field lines given in [Magnetic Fields and Magnetic Field Lines](#) are needed for more detail. There is a simple formula for the **magnetic field strength at the center of a circular loop**. It is

Equation:

$$B = \frac{\mu_0 I}{2R} \text{ (at center of loop),}$$

where R is the radius of the loop. This equation is very similar to that for a straight wire, but it is valid *only* at the center of a circular loop of wire. The similarity of the equations does indicate that similar field strength can be obtained at the center of a loop. One way to get a larger field is to have N loops; then, the field is $B = N\mu_0 I / (2R)$. Note that the larger the loop, the smaller the field at its center, because the current is farther away.

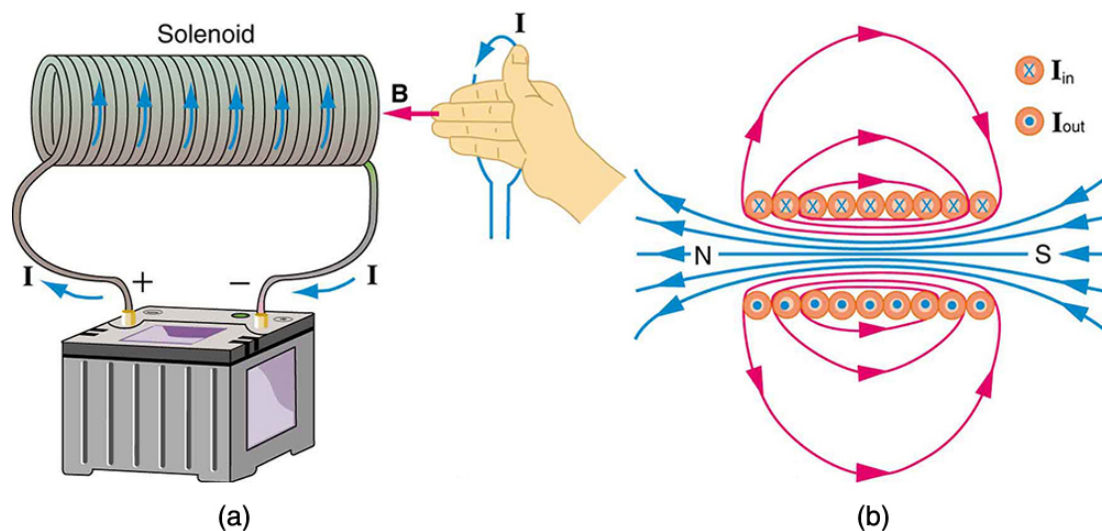


- (a) RHR-2 gives the direction of the magnetic field inside and outside a current-carrying loop. (b) More detailed mapping with compasses or with a

Hall probe completes the picture. The field is similar to that of a bar magnet.

Magnetic Field Produced by a Current-Carrying Solenoid

A **solenoid** is a long coil of wire (with many turns or loops, as opposed to a flat loop). Because of its shape, the field inside a solenoid can be very uniform, and also very strong. The field just outside the coils is nearly zero. [\[link\]](#) shows how the field looks and how its direction is given by RHR-2.



(a) Because of its shape, the field inside a solenoid of length l is remarkably uniform in magnitude and direction, as indicated by the straight and uniformly spaced field lines. The field outside the coils is nearly zero. (b) This cutaway shows the magnetic field generated by the current in the solenoid.

The magnetic field inside of a current-carrying solenoid is very uniform in direction and magnitude. Only near the ends does it begin to weaken and change direction. The field outside has similar complexities to flat loops

and bar magnets, but the **magnetic field strength inside a solenoid** is simply

Equation:

$$B = \mu_0 n I \quad (\text{inside a solenoid}),$$

where n is the number of loops per unit length of the solenoid ($n = N/l$, with N being the number of loops and l the length). Note that B is the field strength anywhere in the uniform region of the interior and not just at the center. Large uniform fields spread over a large volume are possible with solenoids, as [\[link\]](#) implies.

Example:

Calculating Field Strength inside a Solenoid

What is the field inside a 2.00-m-long solenoid that has 2000 loops and carries a 1600-A current?

Strategy

To find the field strength inside a solenoid, we use $B = \mu_0 n I$. First, we note the number of loops per unit length is

Equation:

$$n = \frac{N}{l} = \frac{2000}{2.00 \text{ m}} = 1000 \text{ m}^{-1} = 10 \text{ cm}^{-1}.$$

Solution

Substituting known values gives

Equation:

$$\begin{aligned} B &= \mu_0 n I = (4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}) (1000 \text{ m}^{-1}) (1600 \text{ A}) \\ &= 2.01 \text{ T}. \end{aligned}$$

Discussion

This is a large field strength that could be established over a large-diameter solenoid, such as in medical uses of magnetic resonance imaging (MRI).

The very large current is an indication that the fields of this strength are not

easily achieved, however. Such a large current through 1000 loops squeezed into a meter's length would produce significant heating. Higher currents can be achieved by using superconducting wires, although this is expensive. There is an upper limit to the current, since the superconducting state is disrupted by very large magnetic fields.

There are interesting variations of the flat coil and solenoid. For example, the toroidal coil used to confine the reactive particles in tokamaks is much like a solenoid bent into a circle. The field inside a toroid is very strong but circular. Charged particles travel in circles, following the field lines, and collide with one another, perhaps inducing fusion. But the charged particles do not cross field lines and escape the toroid. A whole range of coil shapes are used to produce all sorts of magnetic field shapes. Adding ferromagnetic materials produces greater field strengths and can have a significant effect on the shape of the field. Ferromagnetic materials tend to trap magnetic fields (the field lines bend into the ferromagnetic material, leaving weaker fields outside it) and are used as shields for devices that are adversely affected by magnetic fields, including the Earth's magnetic field.

Note:

PhET Explorations: Generator

Generate electricity with a bar magnet! Discover the physics behind the phenomena by exploring magnets and how you can use them to make a bulb light.

[Generato](#)

[r](#)

Section Summary

- The strength of the magnetic field created by current in a long straight wire is given by

Equation:

$$B = \frac{\mu_0 I}{2\pi r} (\text{long straight wire}),$$

where I is the current, r is the shortest distance to the wire, and the constant $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ is the permeability of free space.

- The direction of the magnetic field created by a long straight wire is given by right hand rule 2 (RHR-2): *Point the thumb of the right hand in the direction of current, and the fingers curl in the direction of the magnetic field loops* created by it.
- The magnetic field created by current following any path is the sum (or integral) of the fields due to segments along the path (magnitude and direction as for a straight wire), resulting in a general relationship between current and field known as Ampere's law.
- The magnetic field strength at the center of a circular loop is given by

Equation:

$$B = \frac{\mu_0 I}{2R} (\text{at center of loop}),$$

where R is the radius of the loop. This equation becomes $B = \mu_0 nI / (2R)$ for a flat coil of N loops. RHR-2 gives the direction of the field about the loop. A long coil is called a solenoid.

- The magnetic field strength inside a solenoid is

Equation:

$$B = \mu_0 nI \text{ (inside a solenoid),}$$

where n is the number of loops per unit length of the solenoid. The field inside is very uniform in magnitude and direction.

Conceptual Questions

Exercise:

Problem:

Make a drawing and use RHR-2 to find the direction of the magnetic field of a current loop in a motor (such as in [\[link\]](#)). Then show that the direction of the torque on the loop is the same as produced by like poles repelling and unlike poles attracting.

Glossary

right hand rule 2 (RHR-2)

a rule to determine the direction of the magnetic field induced by a current-carrying wire: Point the thumb of the right hand in the direction of current, and the fingers curl in the direction of the magnetic field loops

magnetic field strength (magnitude) produced by a long straight current-carrying wire

defined as $B = \frac{\mu_0 I}{2\pi r}$, where I is the current, r is the shortest distance to the wire, and μ_0 is the permeability of free space

permeability of free space

the measure of the ability of a material, in this case free space, to support a magnetic field; the constant $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$

magnetic field strength at the center of a circular loop

defined as $B = \frac{\mu_0 I}{2R}$ where R is the radius of the loop

solenoid

a thin wire wound into a coil that produces a magnetic field when an electric current is passed through it

magnetic field strength inside a solenoid

defined as $B = \mu_0 n I$ where n is the number of loops per unit length of the solenoid ($n = N/l$, with N being the number of loops and l the length)

Biot-Savart law

a physical law that describes the magnetic field generated by an electric current in terms of a specific equation

Ampere's law

the physical law that states that the magnetic field around an electric current is proportional to the current; each segment of current produces a magnetic field like that of a long straight wire, and the total field of any shape current is the vector sum of the fields due to each segment

Maxwell's equations

a set of four equations that describe electromagnetic phenomena

Magnetic Force between Two Parallel Conductors

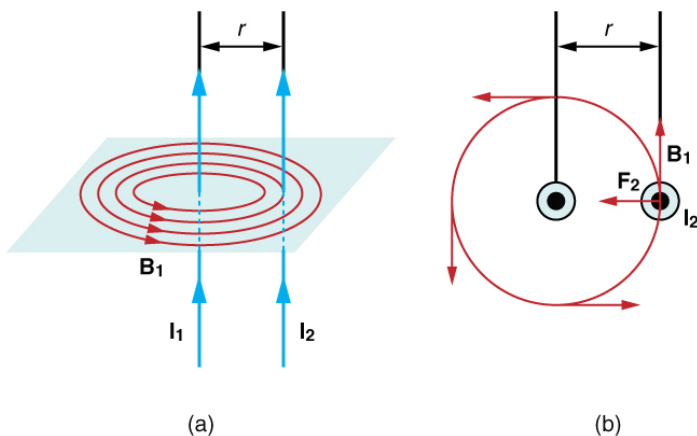
- Describe the effects of the magnetic force between two conductors.
- Calculate the force between two parallel conductors.

You might expect that there are significant forces between current-carrying wires, since ordinary currents produce significant magnetic fields and these fields exert significant forces on ordinary currents. But you might not expect that the force between wires is used to *define* the ampere. It might also surprise you to learn that this force has something to do with why large circuit breakers burn up when they attempt to interrupt large currents.

The force between two long straight and parallel conductors separated by a distance r can be found by applying what we have developed in preceding sections. [\[link\]](#) shows the wires, their currents, the fields they create, and the subsequent forces they exert on one another. Let us consider the field produced by wire 1 and the force it exerts on wire 2 (call the force F_2). The field due to I_1 at a distance r is given to be

Equation:

$$B_1 = \frac{\mu_0 I_1}{2\pi r}.$$



(a) The magnetic field produced by a long straight conductor is perpendicular to a parallel conductor, as indicated by RHR-2. (b) A view

from above of the two wires shown in (a), with one magnetic field line shown for each wire. RHR-1 shows that the force between the parallel conductors is attractive when the currents are in the same direction. A similar analysis shows that the force is repulsive between currents in opposite directions.

This field is uniform along wire 2 and perpendicular to it, and so the force F_2 it exerts on wire 2 is given by $F = IlB \sin \theta$ with $\sin \theta = 1$:

Equation:

$$F_2 = I_2 l B_1.$$

By Newton's third law, the forces on the wires are equal in magnitude, and so we just write F for the magnitude of F_2 . (Note that $F_1 = -F_2$.) Since the wires are very long, it is convenient to think in terms of F/l , the force per unit length. Substituting the expression for B_1 into the last equation and rearranging terms gives

Equation:

$$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}.$$

F/l is the force per unit length between two parallel currents I_1 and I_2 separated by a distance r . The force is attractive if the currents are in the same direction and repulsive if they are in opposite directions.

This force is responsible for the *pinch effect* in electric arcs and plasmas. The force exists whether the currents are in wires or not. In an electric arc, where currents are moving parallel to one another, there is an attraction that squeezes currents into a smaller tube. In large circuit breakers, like those

used in neighborhood power distribution systems, the pinch effect can concentrate an arc between plates of a switch trying to break a large current, burn holes, and even ignite the equipment. Another example of the pinch effect is found in the solar plasma, where jets of ionized material, such as solar flares, are shaped by magnetic forces.

The *operational definition of the ampere* is based on the force between current-carrying wires. Note that for parallel wires separated by 1 meter with each carrying 1 ampere, the force per meter is

Equation:

$$\frac{F}{l} = \frac{(4\pi \times 10^{-7} \text{ T} \cdot \text{m/A})(1 \text{ A})^2}{(2\pi)(1 \text{ m})} = 2 \times 10^{-7} \text{ N/m}.$$

Since μ_0 is exactly $4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ by definition, and because $1 \text{ T} = 1 \text{ N}/(\text{A} \cdot \text{m})$, the force per meter is exactly $2 \times 10^{-7} \text{ N/m}$. This is the basis of the operational definition of the ampere.

Note:

The Ampere

The official definition of the ampere is:

One ampere of current through each of two parallel conductors of infinite length, separated by one meter in empty space free of other magnetic fields, causes a force of exactly $2 \times 10^{-7} \text{ N/m}$ on each conductor.

Infinite-length straight wires are impractical and so, in practice, a current balance is constructed with coils of wire separated by a few centimeters. Force is measured to determine current. This also provides us with a method for measuring the coulomb. We measure the charge that flows for a current of one ampere in one second. That is, $1 \text{ C} = 1 \text{ A} \cdot \text{s}$. For both the ampere and the coulomb, the method of measuring force between conductors is the most accurate in practice.

Section Summary

- The force between two parallel currents I_1 and I_2 , separated by a distance r , has a magnitude per unit length given by

Equation:

$$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}.$$

- The force is attractive if the currents are in the same direction, repulsive if they are in opposite directions.

Conceptual Questions

Exercise:

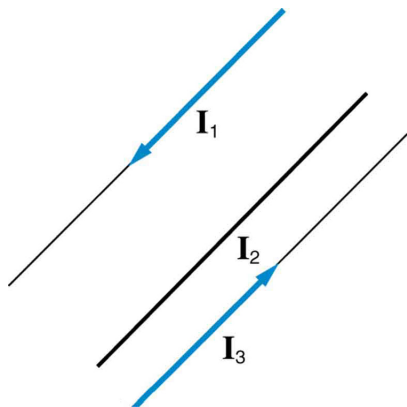
Problem:

Is the force attractive or repulsive between the hot and neutral lines hung from power poles? Why?

Exercise:

Problem:

If you have three parallel wires in the same plane, as in [\[link\]](#), with currents in the outer two running in opposite directions, is it possible for the middle wire to be repelled by both? Attracted by both? Explain.



Three parallel

coplanar wires with
currents in the
outer two in
opposite directions.

Exercise:

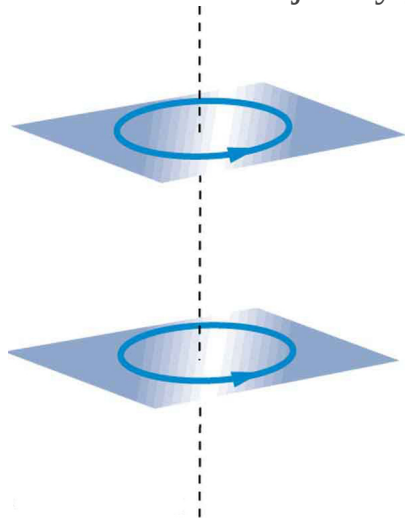
Problem:

Suppose two long straight wires run perpendicular to one another without touching. Does one exert a net force on the other? If so, what is its direction? Does one exert a net torque on the other? If so, what is its direction? Justify your responses by using the right hand rules.

Exercise:

Problem:

Use the right hand rules to show that the force between the two loops in [\[link\]](#) is attractive if the currents are in the same direction and repulsive if they are in opposite directions. Is this consistent with like poles of the loops repelling and unlike poles of the loops attracting? Draw sketches to justify your answers.



Two loops of wire
carrying currents

can exert forces
and torques on one
another.

Exercise:

Problem:

If one of the loops in [\[link\]](#) is tilted slightly relative to the other and their currents are in the same direction, what are the directions of the torques they exert on each other? Does this imply that the poles of the bar magnet-like fields they create will line up with each other if the loops are allowed to rotate?

Exercise:

Problem:

Electric field lines can be shielded by the Faraday cage effect. Can we have magnetic shielding? Can we have gravitational shielding?

Problems & Exercises

Exercise:

Problem:

- (a) The hot and neutral wires supplying DC power to a light-rail commuter train carry 800 A and are separated by 75.0 cm. What is the magnitude and direction of the force between 50.0 m of these wires?
- (b) Discuss the practical consequences of this force, if any.

Solution:

- (a) 8.53 N, repulsive
- (b) This force is repulsive and therefore there is never a risk that the two wires will touch and short circuit.

Exercise:**Problem:**

The force per meter between the two wires of a jumper cable being used to start a stalled car is 0.225 N/m. (a) What is the current in the wires, given they are separated by 2.00 cm? (b) Is the force attractive or repulsive?

Exercise:**Problem:**

A 2.50-m segment of wire supplying current to the motor of a submerged submarine carries 1000 A and feels a 4.00-N repulsive force from a parallel wire 5.00 cm away. What is the direction and magnitude of the current in the other wire?

Solution:

400 A in the opposite direction

Exercise:**Problem:**

The wire carrying 400 A to the motor of a commuter train feels an attractive force of 4.00×10^{-3} N/m due to a parallel wire carrying 5.00 A to a headlight. (a) How far apart are the wires? (b) Are the currents in the same direction?

Exercise:**Problem:**

An AC appliance cord has its hot and neutral wires separated by 3.00 mm and carries a 5.00-A current. (a) What is the average force per meter between the wires in the cord? (b) What is the maximum force per meter between the wires? (c) Are the forces attractive or repulsive? (d) Do appliance cords need any special design features to compensate for these forces?

Solution:

(a) $1.67 \times 10^{-3} \text{ N/m}$

(b) $3.33 \times 10^{-3} \text{ N/m}$

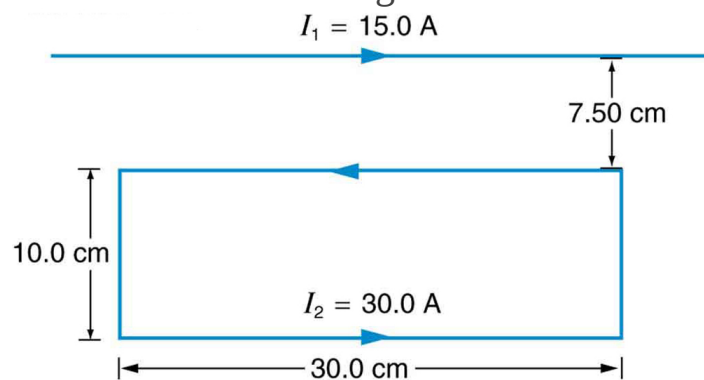
(c) Repulsive

(d) No, these are very small forces

Exercise:

Problem:

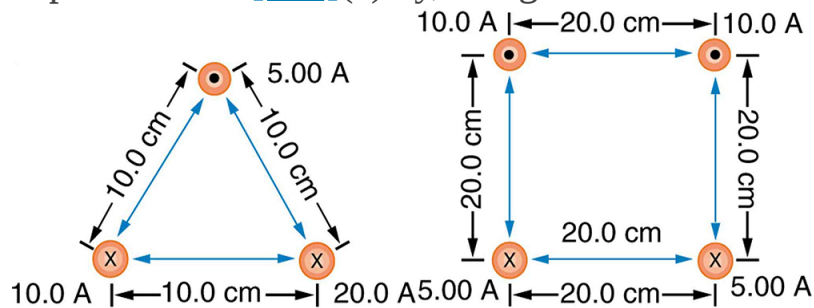
[\[link\]](#) shows a long straight wire near a rectangular current loop. What is the direction and magnitude of the total force on the loop?



Exercise:

Problem:

Find the direction and magnitude of the force that each wire experiences in [\[link\]](#)(a) by, using vector addition.



Solution:

- (a) Top wire: $2.65 \times 10^{-4} \text{ N/m}$, 10.9° to left of up
- (b) Lower left wire: $3.61 \times 10^{-4} \text{ N/m}$, 13.9° down from right
- (c) Lower right wire: $3.46 \times 10^{-4} \text{ N/m}$, 30.0° down from left

Exercise:**Problem:**

Find the direction and magnitude of the force that each wire experiences in [\[link\]](#)(b), using vector addition.

More Applications of Magnetism

- Describe some applications of magnetism.

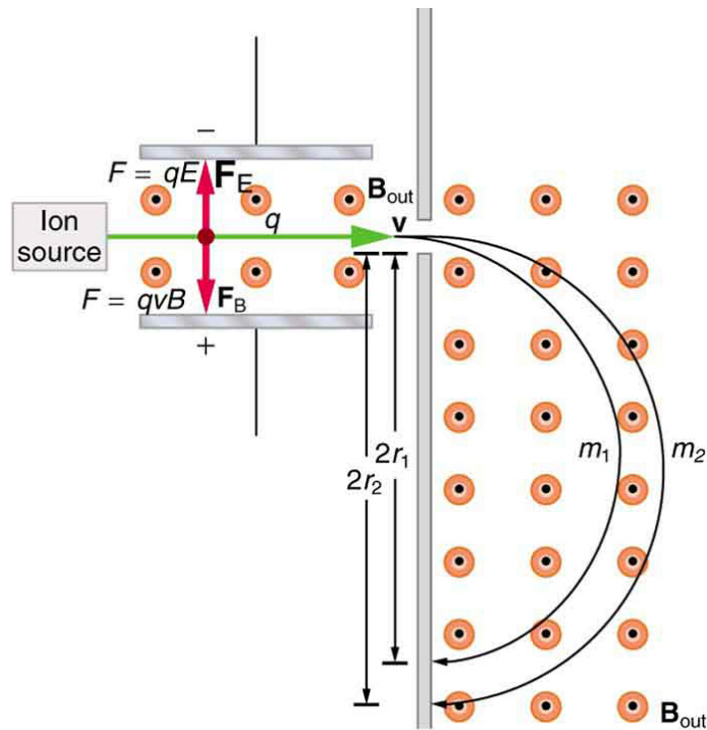
Mass Spectrometry

The curved paths followed by charged particles in magnetic fields can be put to use. A charged particle moving perpendicular to a magnetic field travels in a circular path having a radius r .

Equation:

$$r = \frac{mv}{qB}$$

It was noted that this relationship could be used to measure the mass of charged particles such as ions. A mass spectrometer is a device that measures such masses. Most mass spectrometers use magnetic fields for this purpose, although some of them have extremely sophisticated designs. Since there are five variables in the relationship, there are many possibilities. However, if v , q , and B can be fixed, then the radius of the path r is simply proportional to the mass m of the charged particle. Let us examine one such mass spectrometer that has a relatively simple design. (See [\[link\]](#).) The process begins with an ion source, a device like an electron gun. The ion source gives ions their charge, accelerates them to some velocity v , and directs a beam of them into the next stage of the spectrometer. This next region is a *velocity selector* that only allows particles with a particular value of v to get through.



This mass spectrometer uses a velocity selector to fix v so that the radius of the path is proportional to mass.

The velocity selector has both an electric field and a magnetic field, perpendicular to one another, producing forces in opposite directions on the ions. Only those ions for which the forces balance travel in a straight line into the next region. If the forces balance, then the electric force $F = qE$ equals the magnetic force $F = qvB$, so that $qE = qvB$. Noting that q

Equation:

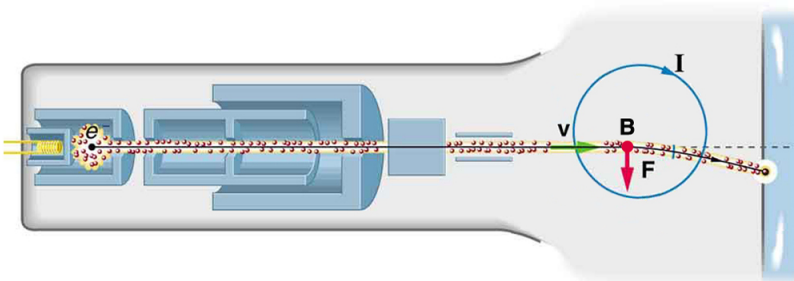
$$v = \frac{E}{B}$$

is the velocity particles must have to make it through the velocity selector, and further, that v can be selected by varying E and B . In the final region, there is only a uniform magnetic field, and so the charged particles move in circular arcs with radii proportional to particle mass. The paths also depend on charge q , but since q is in multiples of electron charges, it is easy to determine and to discriminate between ions in different charge states.

Mass spectrometry today is used extensively in chemistry and biology laboratories to identify chemical and biological substances according to their mass-to-charge ratios. In medicine, mass spectrometers are used to measure the concentration of isotopes used as tracers. Usually, biological molecules such as proteins are very large, so they are broken down into smaller fragments before analyzing. Recently, large virus particles have been analyzed as a whole on mass spectrometers. Sometimes a gas chromatograph or high-performance liquid chromatograph provides an initial separation of the large molecules, which are then input into the mass spectrometer.

Cathode Ray Tubes—CRTs—and the Like

What do non-flat-screen TVs, old computer monitors, x-ray machines, and the 2-mile-long Stanford Linear Accelerator have in common? All of them accelerate electrons, making them different versions of the electron gun. Many of these devices use magnetic fields to steer the accelerated electrons. [\[link\]](#) shows the construction of the type of cathode ray tube (CRT) found in some TVs, oscilloscopes, and old computer monitors. Two pairs of coils are used to steer the electrons, one vertically and the other horizontally, to their desired destination.



The cathode ray tube (CRT) is so named because rays of electrons originate at the cathode in the electron gun. Magnetic coils are used to steer the beam in many CRTs. In this case, the beam is moved down. Another pair of horizontal coils would steer the beam horizontally.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is one of the most useful and rapidly growing medical imaging tools. It non-invasively produces two-dimensional and three-dimensional images of the body that provide important medical information with none of the hazards of x-rays. MRI is based on an effect called **nuclear magnetic resonance (NMR)** in which an externally applied magnetic field interacts with the nuclei of certain atoms, particularly those of hydrogen (protons). These nuclei possess their own small magnetic fields, similar to those of electrons and the current loops discussed earlier in this chapter.

When placed in an external magnetic field, such nuclei experience a torque that pushes or aligns the nuclei into one of two new energy states—depending on the orientation of its spin (analogous to the N pole and S pole in a bar magnet). Transitions from the lower to higher energy state can be achieved by using an external radio frequency signal to “flip” the orientation of the small magnets. (This is actually a quantum mechanical process. The direction of the nuclear magnetic field is quantized as is energy in the radio waves. We will return to these topics in later chapters.) The specific frequency of the radio waves that are absorbed and reemitted depends sensitively on the type of nucleus, the chemical environment, and the external magnetic field strength. Therefore, this is a *resonance* phenomenon in which *nuclei* in a *magnetic* field act like resonators (analogous to those discussed in the treatment of sound in [Oscillatory](#)

[Motion and Waves](#)) that absorb and reemit only certain frequencies. Hence, the phenomenon is named *nuclear magnetic resonance (NMR)*.

NMR has been used for more than 50 years as an analytical tool. It was formulated in 1946 by F. Bloch and E. Purcell, with the 1952 Nobel Prize in Physics going to them for their work. Over the past two decades, NMR has been developed to produce detailed images in a process now called magnetic resonance imaging (MRI), a name coined to avoid the use of the word “nuclear” and the concomitant implication that nuclear radiation is involved. (It is not.) The 2003 Nobel Prize in Medicine went to P. Lauterbur and P. Mansfield for their work with MRI applications.

The largest part of the MRI unit is a superconducting magnet that creates a magnetic field, typically between 1 and 2 T in strength, over a relatively large volume. MRI images can be both highly detailed and informative about structures and organ functions. It is helpful that normal and non-normal tissues respond differently for slight changes in the magnetic field. In most medical images, the protons that are hydrogen nuclei are imaged. (About 2/3 of the atoms in the body are hydrogen.) Their location and density give a variety of medically useful information, such as organ function, the condition of tissue (as in the brain), and the shape of structures, such as vertebral disks and knee-joint surfaces. MRI can also be used to follow the movement of certain ions across membranes, yielding information on active transport, osmosis, dialysis, and other phenomena. With excellent spatial resolution, MRI can provide information about tumors, strokes, shoulder injuries, infections, etc.

An image requires position information as well as the density of a nuclear type (usually protons). By varying the magnetic field slightly over the volume to be imaged, the resonant frequency of the protons is made to vary with position. Broadcast radio frequencies are swept over an appropriate range and nuclei absorb and reemit them only if the nuclei are in a magnetic field with the correct strength. The imaging receiver gathers information through the body almost point by point, building up a tissue map. The reception of reemitted radio waves as a function of frequency thus gives position information. These “slices” or cross sections through the body are only several mm thick. The intensity of the reemitted radio waves is

proportional to the concentration of the nuclear type being flipped, as well as information on the chemical environment in that area of the body.

Various techniques are available for enhancing contrast in images and for obtaining more information. Scans called T1, T2, or proton density scans rely on different relaxation mechanisms of nuclei. Relaxation refers to the time it takes for the protons to return to equilibrium after the external field is turned off. This time depends upon tissue type and status (such as inflammation).

While MRI images are superior to x rays for certain types of tissue and have none of the hazards of x rays, they do not completely supplant x-ray images. MRI is less effective than x rays for detecting breaks in bone, for example, and in imaging breast tissue, so the two diagnostic tools complement each other. MRI images are also expensive compared to simple x-ray images and tend to be used most often where they supply information not readily obtained from x rays. Another disadvantage of MRI is that the patient is totally enclosed with detectors close to the body for about 30 minutes or more, leading to claustrophobia. It is also difficult for the obese patient to be in the magnet tunnel. New “open-MRI” machines are now available in which the magnet does not completely surround the patient.

Over the last decade, the development of much faster scans, called “functional MRI” (fMRI), has allowed us to map the functioning of various regions in the brain responsible for thought and motor control. This technique measures the change in blood flow for activities (thought, experiences, action) in the brain. The nerve cells increase their consumption of oxygen when active. Blood hemoglobin releases oxygen to active nerve cells and has somewhat different magnetic properties when oxygenated than when deoxygenated. With MRI, we can measure this and detect a blood oxygen-dependent signal. Most of the brain scans today use fMRI.

Other Medical Uses of Magnetic Fields

Currents in nerve cells and the heart create magnetic fields like any other currents. These can be measured but with some difficulty since their strengths are about 10^{-6} to 10^{-8} less than the Earth’s magnetic field. Recording of the heart’s magnetic field as it beats is called a

magnetocardiogram (MCG), while measurements of the brain's magnetic field is called a **magnetoencephalogram (MEG)**. Both give information that differs from that obtained by measuring the electric fields of these organs (ECGs and EEGs), but they are not yet of sufficient importance to make these difficult measurements common.

In both of these techniques, the sensors do not touch the body. MCG can be used in fetal studies, and is probably more sensitive than echocardiography. MCG also looks at the heart's electrical activity whose voltage output is too small to be recorded by surface electrodes as in EKG. It has the potential of being a rapid scan for early diagnosis of cardiac ischemia (obstruction of blood flow to the heart) or problems with the fetus.

MEG can be used to identify abnormal electrical discharges in the brain that produce weak magnetic signals. Therefore, it looks at brain activity, not just brain structure. It has been used for studies of Alzheimer's disease and epilepsy. Advances in instrumentation to measure very small magnetic fields have allowed these two techniques to be used more in recent years. What is used is a sensor called a SQUID, for superconducting quantum interference device. This operates at liquid helium temperatures and can measure magnetic fields thousands of times smaller than the Earth's.

Finally, there is a burgeoning market for magnetic cures in which magnets are applied in a variety of ways to the body, from magnetic bracelets to magnetic mattresses. The best that can be said for such practices is that they are apparently harmless, unless the magnets get close to the patient's computer or magnetic storage disks. Claims are made for a broad spectrum of benefits from cleansing the blood to giving the patient more energy, but clinical studies have not verified these claims, nor is there an identifiable mechanism by which such benefits might occur.

Note:

PhET Explorations: Magnet and Compass

Ever wonder how a compass worked to point you to the Arctic? Explore the interactions between a compass and bar magnet, and then add the Earth and find the surprising answer! Vary the magnet's strength, and see how

things change both inside and outside. Use the field meter to measure how the magnetic field changes.

<https://archive.cnx.org/specials/5ca3e2cc-ae74-11e5-b6d3-f3c228f04b5c/magnet-and-compass/#sim-bar-magnet>

Section Summary

- Crossed (perpendicular) electric and magnetic fields act as a velocity filter, giving equal and opposite forces on any charge with velocity perpendicular to the fields and of magnitude

Equation:

$$v = \frac{E}{B}.$$

Conceptual Questions

Exercise:

Problem:

Measurements of the weak and fluctuating magnetic fields associated with brain activity are called magnetoencephalograms (MEGs). Do the brain's magnetic fields imply coordinated or uncoordinated nerve impulses? Explain.

Exercise:

Problem:

Discuss the possibility that a Hall voltage would be generated on the moving heart of a patient during MRI imaging. Also discuss the same effect on the wires of a pacemaker. (The fact that patients with pacemakers are not given MRIs is significant.)

Exercise:

Problem:

A patient in an MRI unit turns his head quickly to one side and experiences momentary dizziness and a strange taste in his mouth. Discuss the possible causes.

Exercise:**Problem:**

You are told that in a certain region there is either a uniform electric or magnetic field. What measurement or observation could you make to determine the type? (Ignore the Earth's magnetic field.)

Exercise:**Problem:**

An example of magnetohydrodynamics (MHD) comes from the flow of a river (salty water). This fluid interacts with the Earth's magnetic field to produce a potential difference between the two river banks. How would you go about calculating the potential difference?

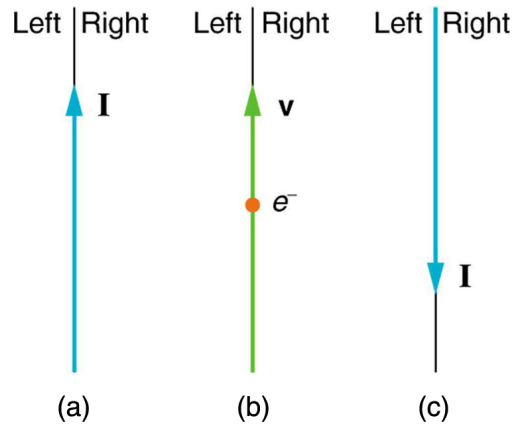
Exercise:**Problem:**

Draw gravitational field lines between 2 masses, electric field lines between a positive and a negative charge, electric field lines between 2 positive charges and magnetic field lines around a magnet. Qualitatively describe the differences between the fields and the entities responsible for the field lines.

Problems & Exercises**Exercise:**

Problem:

Indicate whether the magnetic field created in each of the three situations shown in [\[link\]](#) is into or out of the page on the left and right of the current.

**Solution:**

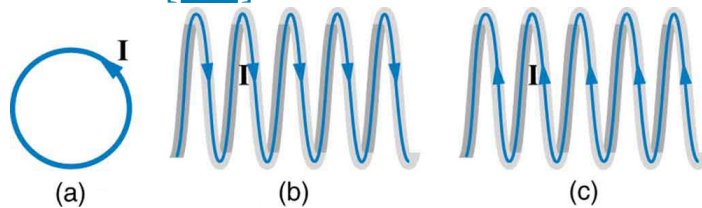
(a) right-into page, left-out of page

(b) right-out of page, left-into page

(c) right-out of page, left-into page

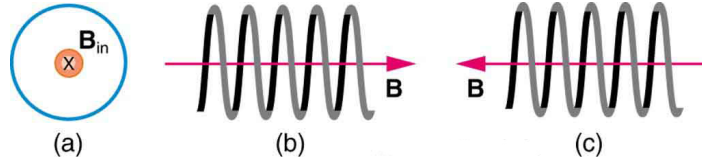
Exercise:**Problem:**

What are the directions of the fields in the center of the loop and coils shown in [\[link\]](#)?

**Exercise:**

Problem:

What are the directions of the currents in the loop and coils shown in [\[link\]](#)?

**Solution:**

- (a) clockwise
- (b) clockwise as seen from the left
- (c) clockwise as seen from the right

Exercise:**Problem:**

To see why an MRI utilizes iron to increase the magnetic field created by a coil, calculate the current needed in a 400-loop-per-meter circular coil 0.660 m in radius to create a 1.20-T field (typical of an MRI instrument) at its center with no iron present. The magnetic field of a proton is approximately like that of a circular current loop 0.650×10^{-15} m in radius carrying 1.05×10^4 A. What is the field at the center of such a loop?

Solution:

$$1.01 \times 10^{13} \text{ T}$$

Exercise:**Problem:**

Inside a motor, 30.0 A passes through a 250-turn circular loop that is 10.0 cm in radius. What is the magnetic field strength created at its center?

Exercise:**Problem:**

Nonnuclear submarines use batteries for power when submerged. (a) Find the magnetic field 50.0 cm from a straight wire carrying 1200 A from the batteries to the drive mechanism of a submarine. (b) What is the field if the wires to and from the drive mechanism are side by side? (c) Discuss the effects this could have for a compass on the submarine that is not shielded.

Solution:

(a) $4.80 \times 10^{-4} \text{ T}$

(b) Zero

(c) If the wires are not paired, the field is about 10 times stronger than Earth's magnetic field and so could severely disrupt the use of a compass.

Exercise:**Problem:**

How strong is the magnetic field inside a solenoid with 10,000 turns per meter that carries 20.0 A?

Exercise:**Problem:**

What current is needed in the solenoid described in [\[link\]](#) to produce a magnetic field 10^4 times the Earth's magnetic field of $5.00 \times 10^{-5} \text{ T}$?

Solution:

39.8 A

Exercise:

Problem:

How far from the starter cable of a car, carrying 150 A, must you be to experience a field less than the Earth's ($5.00 \times 10^{-5} \text{ T}$)? Assume a long straight wire carries the current. (In practice, the body of your car shields the dashboard compass.)

Exercise:**Problem:**

Measurements affect the system being measured, such as the current loop in [\[link\]](#). (a) Estimate the field the loop creates by calculating the field at the center of a circular loop 20.0 cm in diameter carrying 5.00 A. (b) What is the smallest field strength this loop can be used to measure, if its field must alter the measured field by less than 0.0100%?

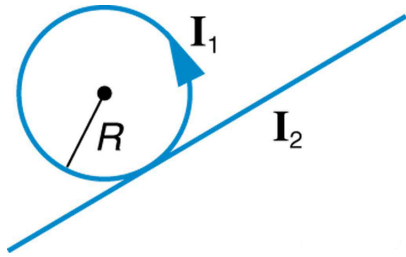
Solution:

(a) $3.14 \times 10^{-5} \text{ T}$

(b) 0.314 T

Exercise:**Problem:**

[\[link\]](#) shows a long straight wire just touching a loop carrying a current I_1 . Both lie in the same plane. (a) What direction must the current I_2 in the straight wire have to create a field at the center of the loop in the direction opposite to that created by the loop? (b) What is the ratio of I_1/I_2 that gives zero field strength at the center of the loop? (c) What is the direction of the field directly above the loop under this circumstance?



Exercise:

Problem:

Find the magnitude and direction of the magnetic field at the point equidistant from the wires in [\[link\]\(a\)](#), using the rules of vector addition to sum the contributions from each wire.

Solution:

$$7.55 \times 10^{-5} \text{ T}, 23.4^\circ$$

Exercise:

Problem:

Find the magnitude and direction of the magnetic field at the point equidistant from the wires in [\[link\]\(b\)](#), using the rules of vector addition to sum the contributions from each wire.

Exercise:

Problem:

What current is needed in the top wire in [\[link\]\(a\)](#) to produce a field of zero at the point equidistant from the wires, if the currents in the bottom two wires are both 10.0 A into the page?

Solution:

$$10.0 \text{ A}$$

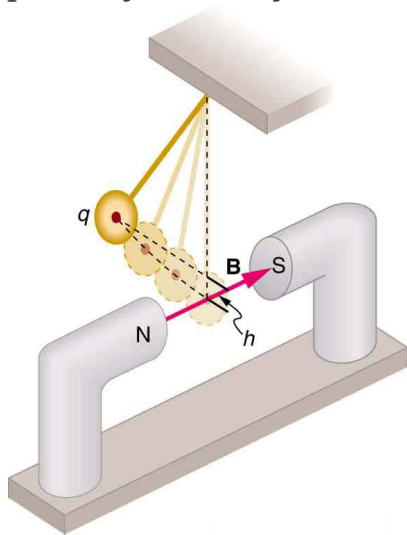
Exercise:

Problem:

Calculate the size of the magnetic field 20 m below a high voltage power line. The line carries 450 MW at a voltage of 300,000 V.

Exercise:**Problem: Integrated Concepts**

(a) A pendulum is set up so that its bob (a thin copper disk) swings between the poles of a permanent magnet as shown in [\[link\]](#). What is the magnitude and direction of the magnetic force on the bob at the lowest point in its path, if it has a positive $0.250\ \mu\text{C}$ charge and is released from a height of 30.0 cm above its lowest point? The magnetic field strength is 1.50 T. (b) What is the acceleration of the bob at the bottom of its swing if its mass is 30.0 grams and it is hung from a flexible string? Be certain to include a free-body diagram as part of your analysis.

**Solution:**

(a) $9.09 \times 10^{-7}\ \text{N}$ upward

(b) $3.03 \times 10^{-5}\ \text{m/s}^2$

Exercise:

Problem: Integrated Concepts

(a) What voltage will accelerate electrons to a speed of $6.00 \times 10^{-7} \text{ m/s}$? (b) Find the radius of curvature of the path of a *proton* accelerated through this potential in a 0.500-T field and compare this with the radius of curvature of an electron accelerated through the same potential.

Exercise:**Problem: Integrated Concepts**

Find the radius of curvature of the path of a 25.0-MeV proton moving perpendicularly to the 1.20-T field of a cyclotron.

Solution:

60.2 cm

Exercise:**Problem: Integrated Concepts**

To construct a nonmechanical water meter, a 0.500-T magnetic field is placed across the supply water pipe to a home and the Hall voltage is recorded. (a) Find the flow rate in liters per second through a 3.00-cm-diameter pipe if the Hall voltage is 60.0 mV. (b) What would the Hall voltage be for the same flow rate through a 10.0-cm-diameter pipe with the same field applied?

Exercise:**Problem: Integrated Concepts**

(a) Using the values given for an MHD drive in [\[link\]](#), and assuming the force is uniformly applied to the fluid, calculate the pressure created in N/m^2 . (b) Is this a significant fraction of an atmosphere?

Solution:

(a) $1.02 \times 10^3 \text{ N/m}^2$

(b) Not a significant fraction of an atmosphere

Exercise:**Problem: Integrated Concepts**

(a) Calculate the maximum torque on a 50-turn, 1.50 cm radius circular current loop carrying 50 μA in a 0.500-T field. (b) If this coil is to be used in a galvanometer that reads 50 μA full scale, what force constant spring must be used, if it is attached 1.00 cm from the axis of rotation and is stretched by the 60° arc moved?

Exercise:**Problem: Integrated Concepts**

A current balance used to define the ampere is designed so that the current through it is constant, as is the distance between wires. Even so, if the wires change length with temperature, the force between them will change. What percent change in force per degree will occur if the wires are copper?

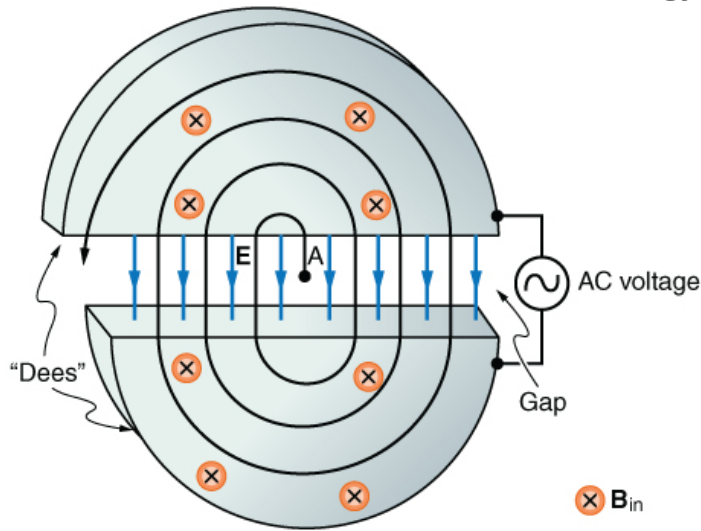
Solution:

$$17.0 \times 10^{-4} \% / ^\circ\text{C}$$

Exercise:**Problem: Integrated Concepts**

(a) Show that the period of the circular orbit of a charged particle moving perpendicularly to a uniform magnetic field is $T = 2\pi m / (qB)$. (b) What is the frequency f ? (c) What is the angular

velocity ω ? Note that these results are independent of the velocity and radius of the orbit and, hence, of the energy of the particle. ([link](#).)



Cyclotrons accelerate charged particles orbiting in a magnetic field by placing an AC voltage on the metal Dees, between which the particles move, so that energy is added twice each orbit. The frequency is constant, since it is independent of the particle energy—the radius of the orbit simply increases with energy until the particles approach the edge and are extracted for various experiments and applications.

Exercise:

Problem: Integrated Concepts

A cyclotron accelerates charged particles as shown in [link](#). Using the results of the previous problem, calculate the frequency of the accelerating voltage needed for a proton in a 1.20-T field.

Solution:

18.3 MHz

Exercise:**Problem: Integrated Concepts**

(a) A 0.140-kg baseball, pitched at 40.0 m/s horizontally and perpendicular to the Earth's horizontal 5.00×10^{-5} T field, has a 100-nC charge on it. What distance is it deflected from its path by the magnetic force, after traveling 30.0 m horizontally? (b) Would you suggest this as a secret technique for a pitcher to throw curve balls?

Exercise:**Problem: Integrated Concepts**

(a) What is the direction of the force on a wire carrying a current due east in a location where the Earth's field is due north? Both are parallel to the ground. (b) Calculate the force per meter if the wire carries 20.0 A and the field strength is 3.00×10^{-5} T. (c) What diameter copper wire would have its weight supported by this force? (d) Calculate the resistance per meter and the voltage per meter needed.

Solution:

(a) Straight up

(b) 6.00×10^{-4} N/m

(c) 94.1 μ m

(d) 2.47 Ω /m, 49.4 V/m

Exercise:**Problem: Integrated Concepts**

One long straight wire is to be held directly above another by repulsion between their currents. The lower wire carries 100 A and the wire 7.50 cm above it is 10-gauge (2.588 mm diameter) copper wire. (a) What current must flow in the upper wire, neglecting the Earth's field? (b) What is the smallest current if the Earth's 3.00×10^{-5} T field is parallel to the ground and is not neglected? (c) Is the supported wire in a stable or unstable equilibrium if displaced vertically? If displaced horizontally?

Exercise:

Problem: Unreasonable Results

(a) Find the charge on a baseball, thrown at 35.0 m/s perpendicular to the Earth's 5.00×10^{-5} T field, that experiences a 1.00-N magnetic force. (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) 571 C

(b) Impossible to have such a large separated charge on such a small object.

(c) The 1.00-N force is much too great to be realistic in the Earth's field.

Exercise:

Problem: Unreasonable Results

A charged particle having mass 6.64×10^{-27} kg (that of a helium atom) moving at 8.70×10^5 m/s perpendicular to a 1.50-T magnetic field travels in a circular path of radius 16.0 mm. (a) What is the charge of the particle? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Unreasonable Results

An inventor wants to generate 120-V power by moving a 1.00-m-long wire perpendicular to the Earth's 5.00×10^{-5} T field. (a) Find the speed with which the wire must move. (b) What is unreasonable about this result? (c) Which assumption is responsible?

Solution:

(a) 2.40×10^6 m/s

(b) The speed is too high to be practical $\leq 1\%$ speed of light

(c) The assumption that you could reasonably generate such a voltage with a single wire in the Earth's field is unreasonable

Exercise:**Problem: Unreasonable Results**

Frustrated by the small Hall voltage obtained in blood flow measurements, a medical physicist decides to increase the applied magnetic field strength to get a 0.500-V output for blood moving at 30.0 cm/s in a 1.50-cm-diameter vessel. (a) What magnetic field strength is needed? (b) What is unreasonable about this result? (c) Which premise is responsible?

Exercise:**Problem: Unreasonable Results**

A surveyor 100 m from a long straight 200-kV DC power line suspects that its magnetic field may equal that of the Earth and affect compass readings. (a) Calculate the current in the wire needed to create a 5.00×10^{-5} T field at this distance. (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) 25.0 kA

(b) This current is unreasonably high. It implies a total power delivery in the line of $50.0 \times 10^9 \text{ W}$, which is much too high for standard transmission lines.

(c) 100 meters is a long distance to obtain the required field strength. Also coaxial cables are used for transmission lines so that there is virtually no field for DC power lines, because of cancellation from opposing currents. The surveyor's concerns are not a problem for his magnetic field measurements.

Exercise:**Problem: Construct Your Own Problem**

Consider a mass separator that applies a magnetic field perpendicular to the velocity of ions and separates the ions based on the radius of curvature of their paths in the field. Construct a problem in which you calculate the magnetic field strength needed to separate two ions that differ in mass, but not charge, and have the same initial velocity. Among the things to consider are the types of ions, the velocities they can be given before entering the magnetic field, and a reasonable value for the radius of curvature of the paths they follow. In addition, calculate the separation distance between the ions at the point where they are detected.

Exercise:**Problem: Construct Your Own Problem**

Consider using the torque on a current-carrying coil in a magnetic field to detect relatively small magnetic fields (less than the field of the Earth, for example). Construct a problem in which you calculate the maximum torque on a current-carrying loop in a magnetic field. Among the things to be considered are the size of the coil, the number

of loops it has, the current you pass through the coil, and the size of the field you wish to detect. Discuss whether the torque produced is large enough to be effectively measured. Your instructor may also wish for you to consider the effects, if any, of the field produced by the coil on the surroundings that could affect detection of the small field.

Glossary

magnetic resonance imaging (MRI)

a medical imaging technique that uses magnetic fields create detailed images of internal tissues and organs

nuclear magnetic resonance (NMR)

a phenomenon in which an externally applied magnetic field interacts with the nuclei of certain atoms

magnetocardiogram (MCG)

a recording of the heart's magnetic field as it beats

magnetoencephalogram (MEG)

a measurement of the brain's magnetic field

Introduction to Electromagnetic Induction, AC Circuits and Electrical Technologies

class="introduction"

These wind turbines in the Thames Estuary in the UK are an example of induction at work.

Wind pushes the blades of the turbine, spinning a shaft attached to magnets.

The magnets spin around a conductive coil, inducing an electric current in the coil, and eventually feeding the electrical grid. (credit: modification of work by Petr Kratochvil)



Nature's displays of symmetry are beautiful and alluring. A butterfly's wings exhibit an appealing symmetry in a complex system. (See [\[link\]](#).) The laws of physics display symmetries at the most basic level—these symmetries are a source of wonder and imply deeper meaning. Since we place a high value on symmetry, we look for it when we explore nature. The remarkable thing is that we find it.



Physics, like this butterfly, has inherent symmetries. (credit: Thomas Bresson)

The hint of symmetry between electricity and magnetism found in the preceding chapter will be elaborated upon in this chapter. Specifically, we know that a current creates a magnetic field. If nature is symmetric here, then perhaps a magnetic field can create a current. The Hall effect is a voltage caused by a magnetic force. That voltage could drive a current. Historically, it was very shortly after Oersted discovered currents cause magnetic fields that other scientists asked the following question: Can magnetic fields cause currents? The answer was soon found by experiment to be yes. In 1831, some 12 years after Oersted's discovery, the English scientist Michael Faraday (1791–1862) and the American scientist Joseph Henry (1797–1878) independently demonstrated that magnetic fields can produce currents. The basic process of generating emfs (electromotive force) and, hence, currents with magnetic fields is known as **induction**; this process is also called magnetic induction to distinguish it from charging by induction, which utilizes the Coulomb force.

Today, currents induced by magnetic fields are essential to our technological society. The ubiquitous generator—found in automobiles, on bicycles, in nuclear power plants, and so on—uses magnetism to generate current. Other devices that use magnetism to induce currents include pickup coils in electric guitars, transformers of every size, certain microphones, airport security gates, and damping mechanisms on sensitive chemical balances. Not so familiar perhaps, but important nevertheless, is that the behavior of AC circuits depends strongly on the effect of magnetic fields on currents.

Glossary

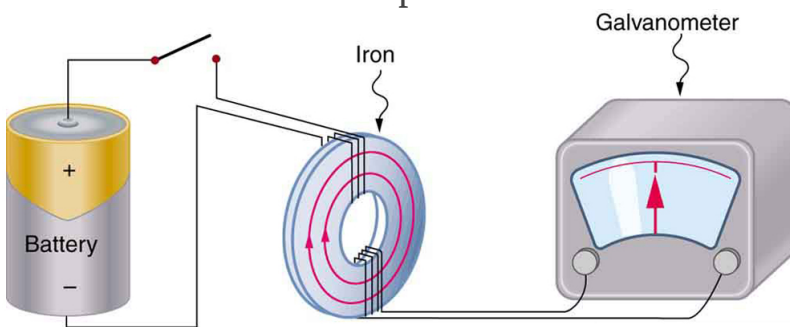
induction

(magnetic induction) the creation of emfs and hence currents by magnetic fields

Induced Emf and Magnetic Flux

- Calculate the flux of a uniform magnetic field through a loop of arbitrary orientation.
- Describe methods to produce an electromotive force (emf) with a magnetic field or magnet and a loop of wire.

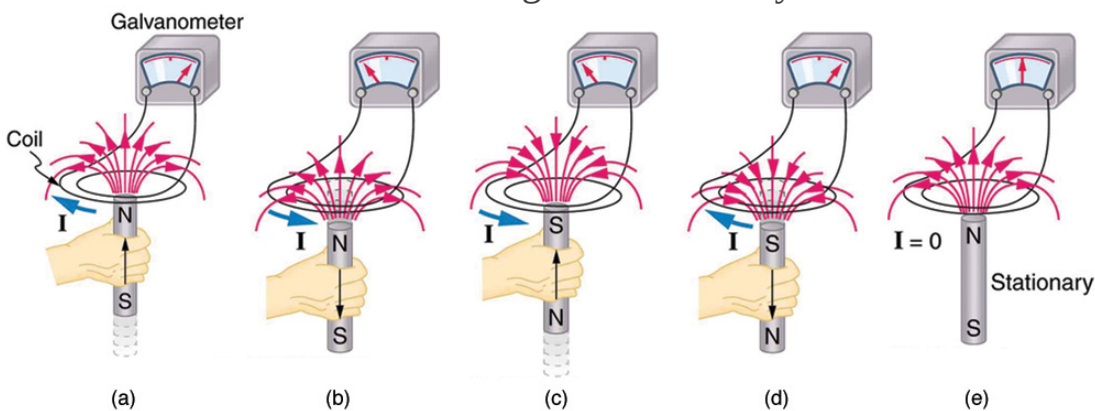
The apparatus used by Faraday to demonstrate that magnetic fields can create currents is illustrated in [\[link\]](#). When the switch is closed, a magnetic field is produced in the coil on the top part of the iron ring and transmitted to the coil on the bottom part of the ring. The galvanometer is used to detect any current induced in the coil on the bottom. It was found that each time the switch is closed, the galvanometer detects a current in one direction in the coil on the bottom. (You can also observe this in a physics lab.) Each time the switch is opened, the galvanometer detects a current in the opposite direction. Interestingly, if the switch remains closed or open for any length of time, there is no current through the galvanometer. *Closing and opening the switch* induces the current. It is the *change* in magnetic field that creates the current. More basic than the current that flows is the emf that causes it. The current is a result of an *emf induced by a changing magnetic field*, whether or not there is a path for current to flow.



Faraday's apparatus for demonstrating that a magnetic field can produce a current. A change in the field produced by the top coil induces an emf and, hence, a current in the bottom coil. When the switch is opened and closed, the galvanometer registers currents in opposite directions. No current flows

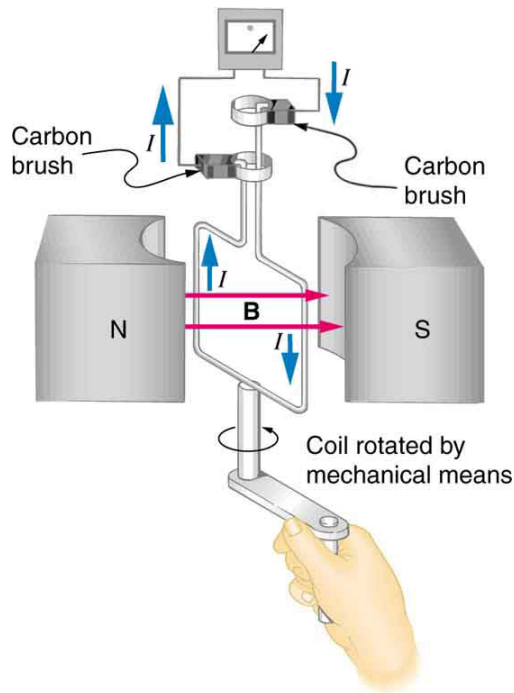
through the galvanometer when the switch remains closed or open.

An experiment easily performed and often done in physics labs is illustrated in [\[link\]](#). An emf is induced in the coil when a bar magnet is pushed in and out of it. Emfs of opposite signs are produced by motion in opposite directions, and the emfs are also reversed by reversing poles. The same results are produced if the coil is moved rather than the magnet—it is the relative motion that is important. The faster the motion, the greater the emf, and there is no emf when the magnet is stationary relative to the coil.



Movement of a magnet relative to a coil produces emfs as shown. The same emfs are produced if the coil is moved relative to the magnet. The greater the speed, the greater the magnitude of the emf, and the emf is zero when there is no motion.

The method of inducing an emf used in most electric generators is shown in [\[link\]](#). A coil is rotated in a magnetic field, producing an alternating current emf, which depends on rotation rate and other factors that will be explored in later sections. Note that the generator is remarkably similar in construction to a motor (another symmetry).



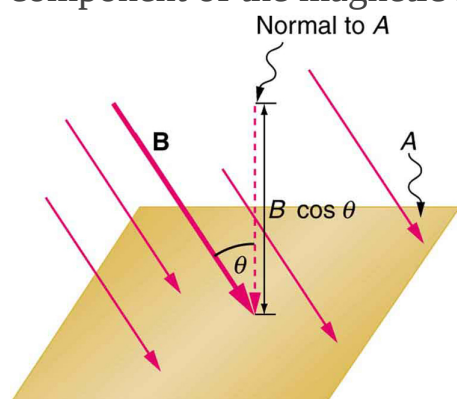
Rotation of a coil in a magnetic field produces an emf. This is the basic construction of a generator, where work done to turn the coil is converted to electric energy. Note the generator is very similar in construction to a motor.

So we see that changing the magnitude or direction of a magnetic field produces an emf. Experiments revealed that there is a crucial quantity called the **magnetic flux**, Φ , given by

Equation:

$$\Phi = BA \cos \theta,$$

where B is the magnetic field strength over an area A , at an angle θ with the perpendicular to the area as shown in [\[link\]](#). **Any change in magnetic flux Φ induces an emf.** This process is defined to be **electromagnetic induction**. Units of magnetic flux Φ are $\text{T} \cdot \text{m}^2$. As seen in [\[link\]](#), $B \cos \theta = B_{\perp}$, which is the component of B perpendicular to the area A . Thus magnetic flux is $\Phi = B_{\perp} A$, the product of the area and the component of the magnetic field perpendicular to it.



$$\Phi = BA \cos \theta = B_{\perp} A$$

Magnetic flux Φ is related to the magnetic field and the area over which it exists. The flux $\Phi = BA \cos \theta$ is related to induction; any change in Φ induces an emf.

All induction, including the examples given so far, arises from some change in magnetic flux Φ . For example, Faraday changed B and hence Φ when opening and closing the switch in his apparatus (shown in [\[link\]](#)). This is also true for the bar magnet and coil shown in [\[link\]](#). When rotating the coil of a generator, the angle θ and, hence, Φ is changed. Just how great an emf and what direction it takes depend on the change in Φ and how rapidly the change is made, as examined in the next section.

Section Summary

- The crucial quantity in induction is magnetic flux Φ , defined to be $\Phi = BA \cos \theta$, where B is the magnetic field strength over an area A at an angle θ with the perpendicular to the area.
- Units of magnetic flux Φ are $\text{T} \cdot \text{m}^2$.
- Any change in magnetic flux Φ induces an emf—the process is defined to be electromagnetic induction.

Conceptual Questions

Exercise:

Problem:

How do the multiple-loop coils and iron ring in the version of Faraday's apparatus shown in [\[link\]](#) enhance the observation of induced emf?

Exercise:

Problem:

When a magnet is thrust into a coil as in [\[link\]](#)(a), what is the direction of the force exerted by the coil on the magnet? Draw a diagram showing the direction of the current induced in the coil and the magnetic field it produces, to justify your response. How does the magnitude of the force depend on the resistance of the galvanometer?

Exercise:

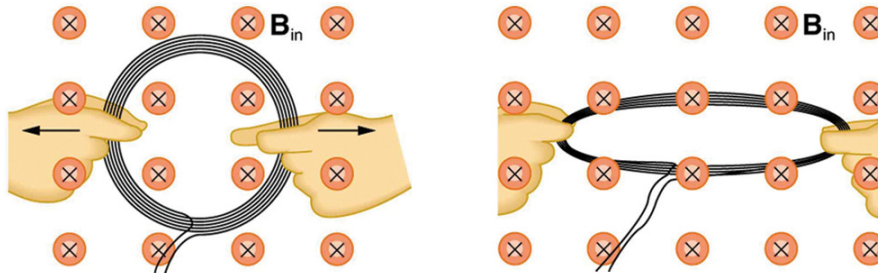
Problem:

Explain how magnetic flux can be zero when the magnetic field is not zero.

Exercise:

Problem:

Is an emf induced in the coil in [\[link\]](#) when it is stretched? If so, state why and give the direction of the induced current.



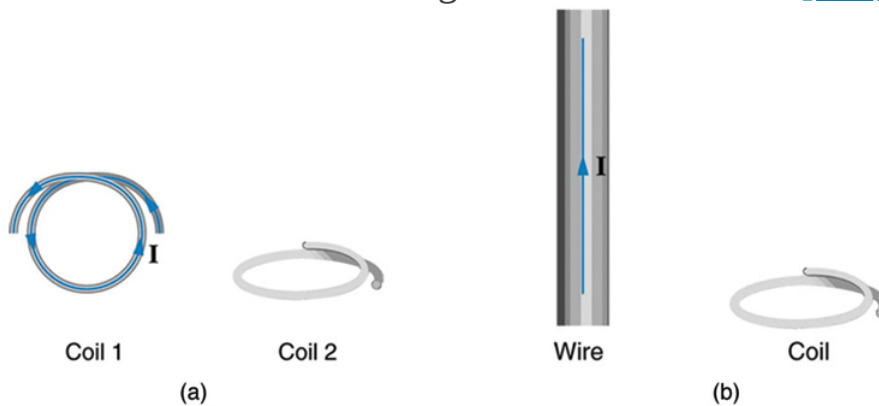
A circular coil of wire is stretched in a magnetic field.

Problems & Exercises

Exercise:

Problem:

What is the value of the magnetic flux at coil 2 in [\[link\]](#) due to coil 1?



(a) The planes of the two coils are perpendicular.

(b) The wire is perpendicular to the plane of the coil.

Solution:

Zero

Exercise:**Problem:**

What is the value of the magnetic flux through the coil in [\[link\]](#)(b) due to the wire?

Glossary

magnetic flux

the amount of magnetic field going through a particular area, calculated with $\Phi = BA \cos \theta$ where B is the magnetic field strength over an area A at an angle θ with the perpendicular to the area

electromagnetic induction

the process of inducing an emf (voltage) with a change in magnetic flux

Faraday's Law of Induction: Lenz's Law

- Calculate emf, current, and magnetic fields using Faraday's Law.
- Explain the physical results of Lenz's Law

Faraday's and Lenz's Law

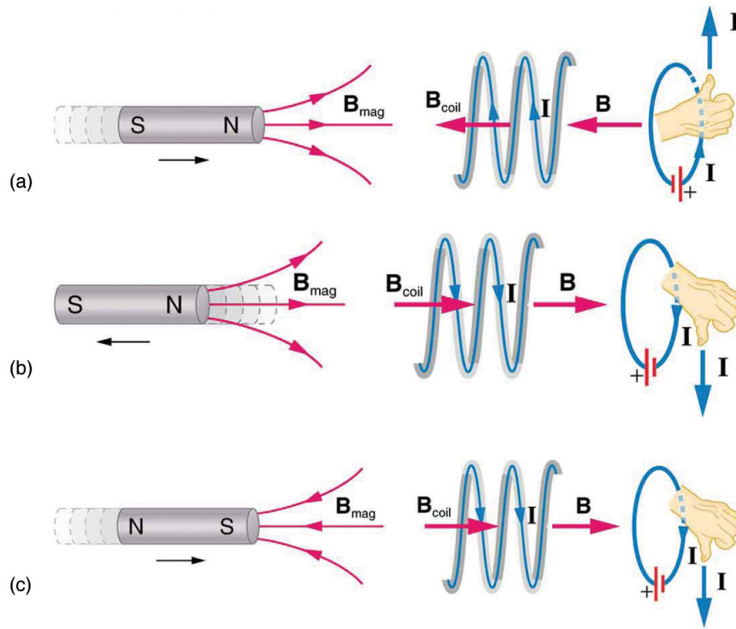
Faraday's experiments showed that the emf induced by a change in magnetic flux depends on only a few factors. First, emf is directly proportional to the change in flux $\Delta\Phi$. Second, emf is greatest when the change in time Δt is smallest—that is, emf is inversely proportional to Δt . Finally, if a coil has N turns, an emf will be produced that is N times greater than for a single coil, so that emf is directly proportional to N . The equation for the emf induced by a change in magnetic flux is

Equation:

$$\text{emf} = -N \frac{\Delta\Phi}{\Delta t}.$$

This relationship is known as **Faraday's law of induction**. The units for emf are volts, as is usual.

The minus sign in Faraday's law of induction is very important. The minus means that *the emf creates a current I and magnetic field B that oppose the change in flux $\Delta\Phi$ —this is known as Lenz's law*. The direction (given by the minus sign) of the emf is so important that it is called **Lenz's law** after the Russian Heinrich Lenz (1804–1865), who, like Faraday and Henry, independently investigated aspects of induction. Faraday was aware of the direction, but Lenz stated it so clearly that he is credited for its discovery. (See [\[link\]](#).)



(a) When this bar magnet is thrust into the coil, the strength of the magnetic field increases in the coil. The current induced in the coil creates another field, in the opposite direction of the bar magnet's to oppose the increase. This is one aspect of *Lenz's law—induction opposes any change in flux*. (b) and (c) are two other situations. Verify for yourself that the direction of the induced B_{coil} shown indeed opposes the change in flux and that the current direction shown is consistent with RHR-2.

Note:

Problem-Solving Strategy for Lenz's Law

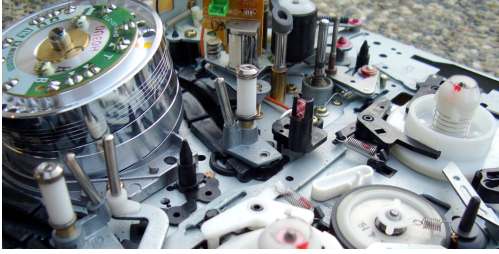
To use Lenz's law to determine the directions of the induced magnetic fields, currents, and emfs:

1. Make a sketch of the situation for use in visualizing and recording directions.
2. Determine the direction of the magnetic field B .
3. Determine whether the flux is increasing or decreasing.
4. Now determine the direction of the induced magnetic field B . It opposes the *change* in flux by adding or subtracting from the original field.
5. Use RHR-2 to determine the direction of the induced current I that is responsible for the induced magnetic field B .
6. The direction (or polarity) of the induced emf will now drive a current in this direction and can be represented as current emerging from the positive terminal of the emf and returning to its negative terminal.

For practice, apply these steps to the situations shown in [\[link\]](#) and to others that are part of the following text material.

Applications of Electromagnetic Induction

There are many applications of Faraday's Law of induction, as we will explore in this chapter and others. At this juncture, let us mention several that have to do with data storage and magnetic fields. A very important application has to do with audio and video *recording tapes*. A plastic tape, coated with iron oxide, moves past a recording head. This recording head is basically a round iron ring about which is wrapped a coil of wire—an electromagnet ([\[link\]](#)). A signal in the form of a varying input current from a microphone or camera goes to the recording head. These signals (which are a function of the signal amplitude and frequency) produce varying magnetic fields at the recording head. As the tape moves past the recording head, the magnetic field orientations of the iron oxide molecules on the tape are changed thus recording the signal. In the playback mode, the magnetized tape is run past another head, similar in structure to the recording head. The different magnetic field orientations of the iron oxide molecules on the tape induces an emf in the coil of wire in the playback head. This signal then is sent to a loudspeaker or video player.



Recording and playback
heads used with audio
and video magnetic tapes.
(credit: Steve Jurvetson)

Similar principles apply to computer hard drives, except at a much faster rate. Here recordings are on a coated, spinning disk. Read heads historically were made to work on the principle of induction. However, the input information is carried in digital rather than analog form – a series of 0's or 1's are written upon the spinning hard drive. Today, most hard drive readout devices do not work on the principle of induction, but use a technique known as *giant magnetoresistance*. (The discovery that weak changes in a magnetic field in a thin film of iron and chromium could bring about much larger changes in electrical resistance was one of the first large successes of nanotechnology.) Another application of induction is found on the magnetic stripe on the back of your personal credit card as used at the grocery store or the ATM machine. This works on the same principle as the audio or video tape mentioned in the last paragraph in which a head reads personal information from your card.

Another application of electromagnetic induction is when electrical signals need to be transmitted across a barrier. Consider the *cochlear implant* shown below. Sound is picked up by a microphone on the outside of the skull and is used to set up a varying magnetic field. A current is induced in a receiver secured in the bone beneath the skin and transmitted to electrodes in the inner ear. Electromagnetic induction can be used in other instances where electric signals need to be conveyed across various media.



Electromagnetic induction used in transmitting electric currents across mediums. The device on the baby's head induces an electrical current in a receiver secured in the bone beneath the skin. (credit: Bjorn Knetsch)

Another contemporary area of research in which electromagnetic induction is being successfully implemented (and with substantial potential) is transcranial magnetic stimulation. A host of disorders, including depression and hallucinations can be traced to irregular localized electrical activity in the brain. In *transcranial magnetic stimulation*, a rapidly varying and very localized magnetic field is placed close to certain sites identified in the brain. Weak electric currents are induced in the identified sites and can result in recovery of electrical functioning in the brain tissue.

Sleep apnea (“the cessation of breath”) affects both adults and infants (especially premature babies and it may be a cause of sudden infant deaths [SID]). In such individuals, breath can stop repeatedly during their sleep. A cessation of more than 20 seconds can be very dangerous. Stroke, heart failure, and tiredness are just some of the possible consequences for a

person having sleep apnea. The concern in infants is the stopping of breath for these longer times. One type of monitor to alert parents when a child is not breathing uses electromagnetic induction. A wire wrapped around the infant's chest has an alternating current running through it. The expansion and contraction of the infant's chest as the infant breathes changes the area through the coil. A pickup coil located nearby has an alternating current induced in it due to the changing magnetic field of the initial wire. If the child stops breathing, there will be a change in the induced current, and so a parent can be alerted.

Note:**Making Connections: Conservation of Energy**

Lenz's law is a manifestation of the conservation of energy. The induced emf produces a current that opposes the change in flux, because a change in flux means a change in energy. Energy can enter or leave, but not instantaneously. Lenz's law is a consequence. As the change begins, the law says induction opposes and, thus, slows the change. In fact, if the induced emf were in the same direction as the change in flux, there would be a positive feedback that would give us free energy from no apparent source—conservation of energy would be violated.

Example:**Calculating Emf: How Great Is the Induced Emf?**

Calculate the magnitude of the induced emf when the magnet in [\[link\]](#)(a) is thrust into the coil, given the following information: the single loop coil has a radius of 6.00 cm and the average value of $B \cos \theta$ (this is given, since the bar magnet's field is complex) increases from 0.0500 T to 0.250 T in 0.100 s.

Strategy

To find the *magnitude* of emf, we use Faraday's law of induction as stated by $\text{emf} = -N \frac{\Delta \Phi}{\Delta t}$, but without the minus sign that indicates direction:

Equation:

$$\text{emf} = N \frac{\Delta\Phi}{\Delta t}.$$

Solution

We are given that $N = 1$ and $\Delta t = 0.100$ s, but we must determine the change in flux $\Delta\Phi$ before we can find emf. Since the area of the loop is fixed, we see that

Equation:

$$\Delta\Phi = \Delta(BA \cos \theta) = A\Delta(B \cos \theta).$$

Now $\Delta(B \cos \theta) = 0.200$ T, since it was given that $B \cos \theta$ changes from 0.0500 to 0.250 T. The area of the loop is

$A = \pi r^2 = (3.14...)(0.060 \text{ m})^2 = 1.13 \times 10^{-2} \text{ m}^2$. Thus,

Equation:

$$\Delta\Phi = (1.13 \times 10^{-2} \text{ m}^2)(0.200 \text{ T}).$$

Entering the determined values into the expression for emf gives

Equation:

$$\text{Emf} = N \frac{\Delta\Phi}{\Delta t} = \frac{(1.13 \times 10^{-2} \text{ m}^2)(0.200 \text{ T})}{0.100 \text{ s}} = 22.6 \text{ mV}.$$

Discussion

While this is an easily measured voltage, it is certainly not large enough for most practical applications. More loops in the coil, a stronger magnet, and faster movement make induction the practical source of voltages that it is.

Note:

PhET Explorations: Faraday's Electromagnetic Lab

Play with a bar magnet and coils to learn about Faraday's law. Move a bar magnet near one or two coils to make a light bulb glow. View the magnetic field lines. A meter shows the direction and magnitude of the current. View the magnetic field lines or use a meter to show the direction and magnitude of the current. You can also play with electromagnets, generators and

transformers!

<https://archive.cnx.org/specials/70b14c10-ae73-11e5-8eb2-b7fbe0c5c7a4/faraday/#sim-bar-magnet>

Section Summary

- Faraday's law of induction states that the emf induced by a change in magnetic flux is

Equation:

$$\text{emf} = -N \frac{\Delta\Phi}{\Delta t}$$

when flux changes by $\Delta\Phi$ in a time Δt .

- If emf is induced in a coil, N is its number of turns.
- The minus sign means that the emf creates a current I and magnetic field B that *oppose the change in flux* $\Delta\Phi$ —this opposition is known as Lenz's law.

Conceptual Questions

Exercise:

Problem:

A person who works with large magnets sometimes places her head inside a strong field. She reports feeling dizzy as she quickly turns her head. How might this be associated with induction?

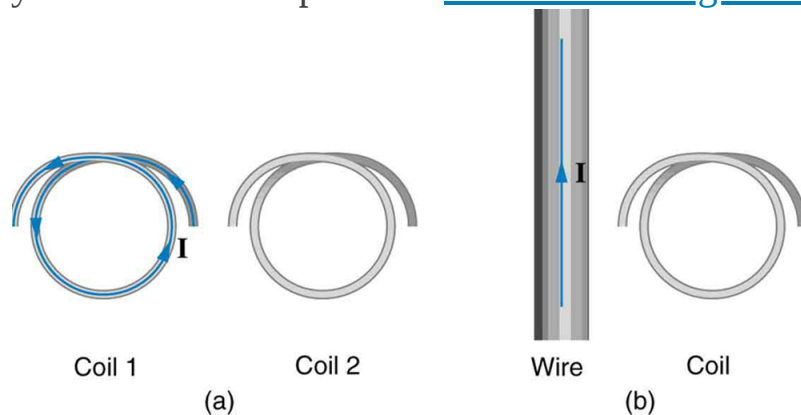
Exercise:

Problem:

A particle accelerator sends high-velocity charged particles down an evacuated pipe. Explain how a coil of wire wrapped around the pipe could detect the passage of individual particles. Sketch a graph of the voltage output of the coil as a single particle passes through it.

Problems & Exercises**Exercise:****Problem:**

Referring to [\[link\]](#)(a), what is the direction of the current induced in coil 2: (a) If the current in coil 1 increases? (b) If the current in coil 1 decreases? (c) If the current in coil 1 is constant? Explicitly show how you follow the steps in the [Problem-Solving Strategy for Lenz's Law](#).



(a) The coils lie in the same plane. (b) The wire is in the plane of the coil

Solution:

(a) CCW

(b) CW

(c) No current induced

Exercise:

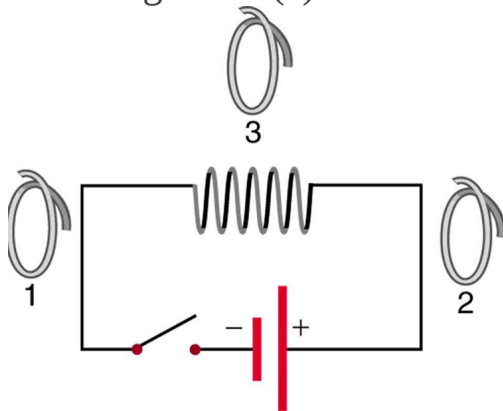
Problem:

Referring to [\[link\]](#)(b), what is the direction of the current induced in the coil: (a) If the current in the wire increases? (b) If the current in the wire decreases? (c) If the current in the wire suddenly changes direction? Explicitly show how you follow the steps in the [Problem-Solving Strategy for Lenz's Law](#).

Exercise:

Problem:

Referring to [\[link\]](#), what are the directions of the currents in coils 1, 2, and 3 (assume that the coils are lying in the plane of the circuit): (a) When the switch is first closed? (b) When the switch has been closed for a long time? (c) Just after the switch is opened?



Solution:

(a) 1 CCW, 2 CCW, 3 CW

(b) 1, 2, and 3 no current induced

(c) 1 CW, 2 CW, 3 CCW

Exercise:

Problem: Repeat the previous problem with the battery reversed.

Exercise:

Problem:

Verify that the units of $\Delta\Phi/\Delta t$ are volts. That is, show that $1 \text{ T} \cdot \text{m}^2/\text{s} = 1 \text{ V}$.

Exercise:

Problem:

Suppose a 50-turn coil lies in the plane of the page in a uniform magnetic field that is directed into the page. The coil originally has an area of 0.250 m^2 . It is stretched to have no area in 0.100 s . What is the direction and magnitude of the induced emf if the uniform magnetic field has a strength of 1.50 T ?

Exercise:

Problem:

(a) An MRI technician moves his hand from a region of very low magnetic field strength into an MRI scanner's 2.00 T field with his fingers pointing in the direction of the field. Find the average emf induced in his wedding ring, given its diameter is 2.20 cm and assuming it takes 0.250 s to move it into the field. (b) Discuss whether this current would significantly change the temperature of the ring.

Solution:

(a) 3.04 mV

(b) As a lower limit on the ring, estimate $R = 1.00 \text{ m}\Omega$. The heat transferred will be 2.31 mJ . This is not a significant amount of heat.

Exercise:

Problem: Integrated Concepts

Referring to the situation in the previous problem: (a) What current is induced in the ring if its resistance is $0.0100\ \Omega$? (b) What average power is dissipated? (c) What magnetic field is induced at the center of the ring? (d) What is the direction of the induced magnetic field relative to the MRI's field?

Exercise:

Problem:

An emf is induced by rotating a 1000-turn, 20.0 cm diameter coil in the Earth's $5.00 \times 10^{-5}\ \text{T}$ magnetic field. What average emf is induced, given the plane of the coil is originally perpendicular to the Earth's field and is rotated to be parallel to the field in 10.0 ms?

Solution:

0.157 V

Exercise:

Problem:

A 0.250 m radius, 500-turn coil is rotated one-fourth of a revolution in 4.17 ms, originally having its plane perpendicular to a uniform magnetic field. (This is 60 rev/s.) Find the magnetic field strength needed to induce an average emf of 10,000 V.

Exercise:

Problem: Integrated Concepts

Approximately how does the emf induced in the loop in [\[link\]](#)(b) depend on the distance of the center of the loop from the wire?

Solution:

proportional to $\frac{1}{r}$

Exercise:

Problem: Integrated Concepts

- (a) A lightning bolt produces a rapidly varying magnetic field. If the bolt strikes the earth vertically and acts like a current in a long straight wire, it will induce a voltage in a loop aligned like that in [\[link\]](#)(b). What voltage is induced in a 1.00 m diameter loop 50.0 m from a 2.00×10^6 A lightning strike, if the current falls to zero in 25.0 μ s?
- (b) Discuss circumstances under which such a voltage would produce noticeable consequences.

Glossary

Faraday's law of induction

the means of calculating the emf in a coil due to changing magnetic flux, given by $\text{emf} = -N \frac{\Delta\Phi}{\Delta t}$

Lenz's law

the minus sign in Faraday's law, signifying that the emf induced in a coil opposes the change in magnetic flux

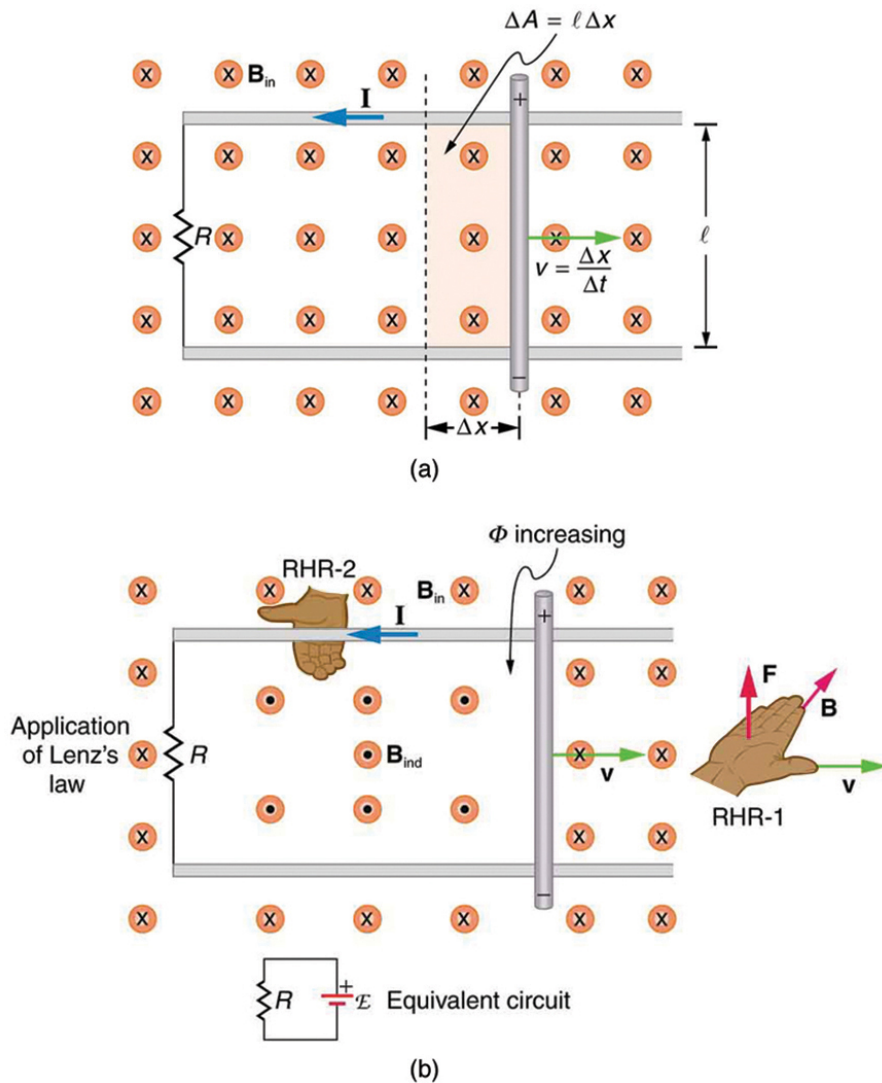
Motional Emf

- Calculate emf, force, magnetic field, and work due to the motion of an object in a magnetic field.

As we have seen, any change in magnetic flux induces an emf opposing that change—a process known as induction. Motion is one of the major causes of induction. For example, a magnet moved toward a coil induces an emf, and a coil moved toward a magnet produces a similar emf. In this section, we concentrate on motion in a magnetic field that is stationary relative to the Earth, producing what is loosely called *motional emf*.

One situation where motional emf occurs is known as the Hall effect and has already been examined. Charges moving in a magnetic field experience the magnetic force $F = qvB \sin \theta$, which moves opposite charges in opposite directions and produces an $\text{emf} = B\ell v$. We saw that the Hall effect has applications, including measurements of B and v . We will now see that the Hall effect is one aspect of the broader phenomenon of induction, and we will find that motional emf can be used as a power source.

Consider the situation shown in [\[link\]](#). A rod is moved at a speed v along a pair of conducting rails separated by a distance ℓ in a uniform magnetic field B . The rails are stationary relative to B and are connected to a stationary resistor R . The resistor could be anything from a light bulb to a voltmeter. Consider the area enclosed by the moving rod, rails, and resistor. B is perpendicular to this area, and the area is increasing as the rod moves. Thus the magnetic flux enclosed by the rails, rod, and resistor is increasing. When flux changes, an emf is induced according to Faraday's law of induction.



(a) A motional emf $= B\ell v$ is induced between the rails when this rod moves to the right in the uniform magnetic field. The magnetic field B is into the page, perpendicular to the moving rod and rails and, hence, to the area enclosed by them. (b) Lenz's law gives the directions of the induced field and current, and the polarity of the induced emf. Since the flux is increasing, the induced field is in the opposite direction, or out of the page. RHR-2 gives the current direction shown, and the polarity of the rod will drive such a current. RHR-1 also indicates the same polarity for the rod. (Note that

the script E symbol used in the equivalent circuit at the bottom of part (b) represents emf.)

To find the magnitude of emf induced along the moving rod, we use Faraday's law of induction without the sign:

Equation:

$$\text{emf} = N \frac{\Delta \Phi}{\Delta t}.$$

Here and below, “emf” implies the magnitude of the emf. In this equation, $N = 1$ and the flux $\Phi = BA \cos \theta$. We have $\theta = 0^\circ$ and $\cos \theta = 1$, since B is perpendicular to A . Now $\Delta \Phi = \Delta(BA) = B\Delta A$, since B is uniform. Note that the area swept out by the rod is $\Delta A = \ell \Delta x$. Entering these quantities into the expression for emf yields

Equation:

$$\text{emf} = \frac{B\Delta A}{\Delta t} = B \frac{\ell \Delta x}{\Delta t}.$$

Finally, note that $\Delta x / \Delta t = v$, the velocity of the rod. Entering this into the last expression shows that

Equation:

$$\text{emf} = B\ell v \quad (B, \ell, \text{ and } v \text{ perpendicular})$$

is the motional emf. This is the same expression given for the Hall effect previously.

Note:

Making Connections: Unification of Forces

There are many connections between the electric force and the magnetic force. The fact that a moving electric field produces a magnetic field and, conversely, a moving magnetic field produces an electric field is part of why electric and magnetic forces are now considered to be different manifestations of the same force. This classic unification of electric and magnetic forces into what is called the electromagnetic force is the inspiration for contemporary efforts to unify other basic forces.

To find the direction of the induced field, the direction of the current, and the polarity of the induced emf, we apply Lenz's law as explained in [Faraday's Law of Induction: Lenz's Law](#). (See [\[link\]](#)(b).) Flux is increasing, since the area enclosed is increasing. Thus the induced field must oppose the existing one and be out of the page. And so the RHR-2 requires that I be counterclockwise, which in turn means the top of the rod is positive as shown.

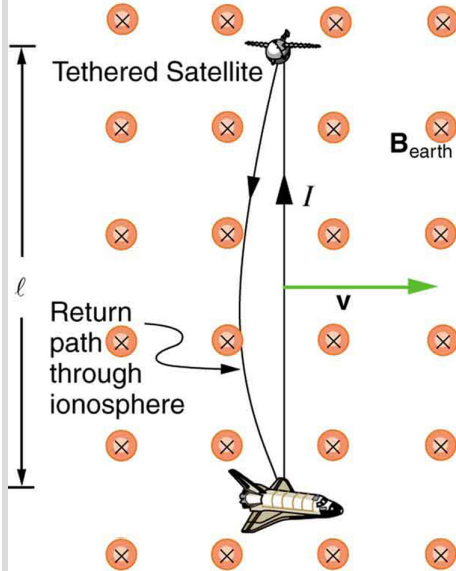
Motional emf also occurs if the magnetic field moves and the rod (or other object) is stationary relative to the Earth (or some observer). We have seen an example of this in the situation where a moving magnet induces an emf in a stationary coil. It is the relative motion that is important. What is emerging in these observations is a connection between magnetic and electric fields. A moving magnetic field produces an electric field through its induced emf. We already have seen that a moving electric field produces a magnetic field—moving charge implies moving electric field and moving charge produces a magnetic field.

Motional emfs in the Earth's weak magnetic field are not ordinarily very large, or we would notice voltage along metal rods, such as a screwdriver, during ordinary motions. For example, a simple calculation of the motional emf of a 1 m rod moving at 3.0 m/s perpendicular to the Earth's field gives $\text{emf} = B\ell v = (5.0 \times 10^{-5} \text{ T})(1.0 \text{ m})(3.0 \text{ m/s}) = 150 \text{ } \mu\text{V}$. This small value is consistent with experience. There is a spectacular exception, however. In 1992 and 1996, attempts were made with the space shuttle to create large motional emfs. The Tethered Satellite was to be let out on a 20 km length of wire as shown in [\[link\]](#), to create a 5 kV emf by moving at

orbital speed through the Earth's field. This emf could be used to convert some of the shuttle's kinetic and potential energy into electrical energy if a complete circuit could be made. To complete the circuit, the stationary ionosphere was to supply a return path for the current to flow. (The ionosphere is the rarefied and partially ionized atmosphere at orbital altitudes. It conducts because of the ionization. The ionosphere serves the same function as the stationary rails and connecting resistor in [\[link\]](#), without which there would not be a complete circuit.) Drag on the current in the cable due to the magnetic force $F = I\ell B \sin \theta$ does the work that reduces the shuttle's kinetic and potential energy and allows it to be converted to electrical energy. The tests were both unsuccessful. In the first, the cable hung up and could only be extended a couple of hundred meters; in the second, the cable broke when almost fully extended. [\[link\]](#) indicates feasibility in principle.

Example:

Calculating the Large Motional Emf of an Object in Orbit



Motional emf as electrical power conversion for the space shuttle is the motivation for the

Tethered Satellite experiment. A 5 kV emf was predicted to be induced in the 20 km long tether while moving at orbital speed in the Earth's magnetic field. The circuit is completed by a return path through the stationary ionosphere.

Calculate the motional emf induced along a 20.0 km long conductor moving at an orbital speed of 7.80 km/s perpendicular to the Earth's 5.00×10^{-5} T magnetic field.

Strategy

This is a straightforward application of the expression for motional emf— $\text{emf} = B\ell v$.

Solution

Entering the given values into $\text{emf} = B\ell v$ gives

Equation:

$$\begin{aligned}\text{emf} &= B\ell v \\ &= (5.00 \times 10^{-5} \text{ T})(2.0 \times 10^4 \text{ m})(7.80 \times 10^3 \text{ m/s}) \\ &= 7.80 \times 10^3 \text{ V}.\end{aligned}$$

Discussion

The value obtained is greater than the 5 kV measured voltage for the shuttle experiment, since the actual orbital motion of the tether is not perpendicular to the Earth's field. The 7.80 kV value is the maximum emf obtained when $\theta = 90^\circ$ and $\sin \theta = 1$.

Section Summary

- An emf induced by motion relative to a magnetic field B is called a *motional emf* and is given by

Equation:

$$\text{emf} = B\ell v \quad (B, \ell, \text{ and } v \text{ perpendicular}),$$

where ℓ is the length of the object moving at speed v relative to the field.

Conceptual Questions

Exercise:

Problem:

Why must part of the circuit be moving relative to other parts, to have usable motional emf? Consider, for example, that the rails in [\[link\]](#) are stationary relative to the magnetic field, while the rod moves.

Exercise:

Problem:

A powerful induction cannon can be made by placing a metal cylinder inside a solenoid coil. The cylinder is forcefully expelled when solenoid current is turned on rapidly. Use Faraday's and Lenz's laws to explain how this works. Why might the cylinder get live/hot when the cannon is fired?

Exercise:

Problem:

An induction stove heats a pot with a coil carrying an alternating current located beneath the pot (and without a hot surface). Can the stove surface be a conductor? Why won't a coil carrying a direct current work?

Exercise:

Problem:

Explain how you could thaw out a frozen water pipe by wrapping a coil carrying an alternating current around it. Does it matter whether or not the pipe is a conductor? Explain.

Problems & Exercises**Exercise:****Problem:**

Use Faraday's law, Lenz's law, and RHR-1 to show that the magnetic force on the current in the moving rod in [\[link\]](#) is in the opposite direction of its velocity.

Exercise:**Problem:**

If a current flows in the Satellite Tether shown in [\[link\]](#), use Faraday's law, Lenz's law, and RHR-1 to show that there is a magnetic force on the tether in the direction opposite to its velocity.

Exercise:**Problem:**

(a) A jet airplane with a 75.0 m wingspan is flying at 280 m/s. What emf is induced between wing tips if the vertical component of the Earth's field is $3.00 \times 10^{-5} \text{ T}$? (b) Is an emf of this magnitude likely to have any consequences? Explain.

Solution:

(a) 0.630 V

(b) No, this is a very small emf.

Exercise:

Problem:

(a) A nonferrous screwdriver is being used in a 2.00 T magnetic field. What maximum emf can be induced along its 12.0 cm length when it moves at 6.00 m/s? (b) Is it likely that this emf will have any consequences or even be noticed?

Exercise:**Problem:**

At what speed must the sliding rod in [\[link\]](#) move to produce an emf of 1.00 V in a 1.50 T field, given the rod's length is 30.0 cm?

Solution:

2.22 m/s

Exercise:**Problem:**

The 12.0 cm long rod in [\[link\]](#) moves at 4.00 m/s. What is the strength of the magnetic field if a 95.0 V emf is induced?

Exercise:**Problem:**

Prove that when B , ℓ , and v are not mutually perpendicular, motional emf is given by $\text{emf} = B\ell v \sin \theta$. If v is perpendicular to B , then θ is the angle between ℓ and B . If ℓ is perpendicular to B , then θ is the angle between v and B .

Exercise:

Problem:

In the August 1992 space shuttle flight, only 250 m of the conducting tether considered in [\[link\]](#) could be let out. A 40.0 V motional emf was generated in the Earth's 5.00×10^{-5} T field, while moving at 7.80×10^3 m/s. What was the angle between the shuttle's velocity and the Earth's field, assuming the conductor was perpendicular to the field?

Exercise:**Problem: Integrated Concepts**

Derive an expression for the current in a system like that in [\[link\]](#), under the following conditions. The resistance between the rails is R , the rails and the moving rod are identical in cross section A and have the same resistivity ρ . The distance between the rails is l , and the rod moves at constant speed v perpendicular to the uniform field B . At time zero, the moving rod is next to the resistance R .

Exercise:**Problem: Integrated Concepts**

The Tethered Satellite in [\[link\]](#) has a mass of 525 kg and is at the end of a 20.0 km long, 2.50 mm diameter cable with the tensile strength of steel. (a) How much does the cable stretch if a 100 N force is exerted to pull the satellite in? (Assume the satellite and shuttle are at the same altitude above the Earth.) (b) What is the effective force constant of the cable? (c) How much energy is stored in it when stretched by the 100 N force?

Exercise:**Problem: Integrated Concepts**

The Tethered Satellite discussed in this module is producing 5.00 kV, and a current of 10.0 A flows. (a) What magnetic drag force does this

produce if the system is moving at 7.80 km/s? (b) How much kinetic energy is removed from the system in 1.00 h, neglecting any change in altitude or velocity during that time? (c) What is the change in velocity if the mass of the system is 100,000 kg? (d) Discuss the long term consequences (say, a week-long mission) on the space shuttle's orbit, noting what effect a decrease in velocity has and assessing the magnitude of the effect.

Solution:

(a) 10.0 N

(b) 2.81×10^8 J

(c) 0.36 m/s

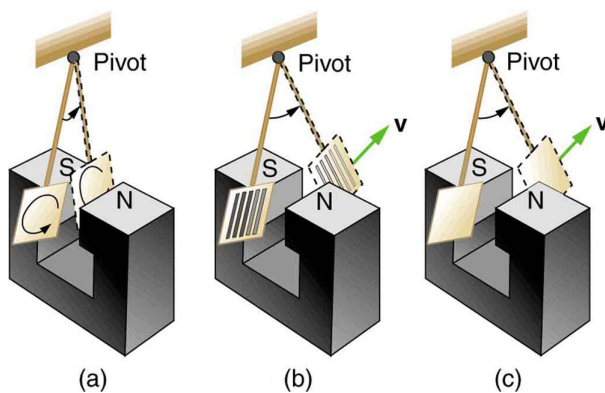
(d) For a week-long mission (168 hours), the change in velocity will be 60 m/s, or approximately 1%. In general, a decrease in velocity would cause the orbit to start spiraling inward because the velocity would no longer be sufficient to keep the circular orbit. The long-term consequences are that the shuttle would require a little more fuel to maintain the desired speed, otherwise the orbit would spiral slightly inward.

Eddy Currents and Magnetic Damping

- Explain the magnitude and direction of an induced eddy current, and the effect this will have on the object it is induced in.
- Describe several applications of magnetic damping.

Eddy Currents and Magnetic Damping

As discussed in [Motional Emf](#), motional emf is induced when a conductor moves in a magnetic field or when a magnetic field moves relative to a conductor. If motional emf can cause a current loop in the conductor, we refer to that current as an **eddy current**. Eddy currents can produce significant drag, called **magnetic damping**, on the motion involved. Consider the apparatus shown in [\[link\]](#), which swings a pendulum bob between the poles of a strong magnet. (This is another favorite physics lab activity.) If the bob is metal, there is significant drag on the bob as it enters and leaves the field, quickly damping the motion. If, however, the bob is a slotted metal plate, as shown in [\[link\]](#)(b), there is a much smaller effect due to the magnet. There is no discernible effect on a bob made of an insulator. Why is there drag in both directions, and are there any uses for magnetic drag?

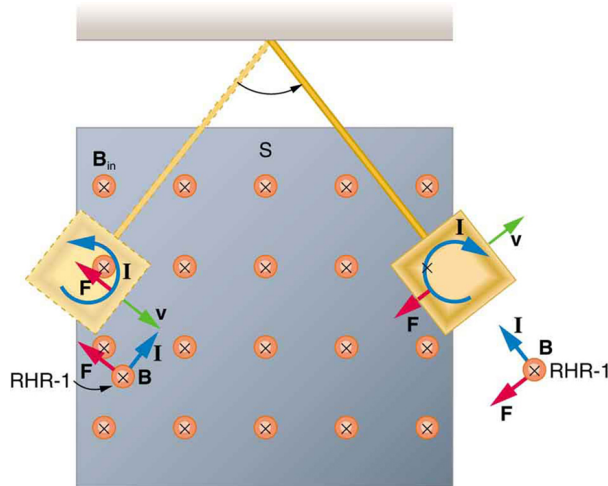


A common physics demonstration device for exploring eddy currents and magnetic damping. (a) The

motion of a metal pendulum bob swinging between the poles of a magnet is quickly damped by the action of eddy currents.

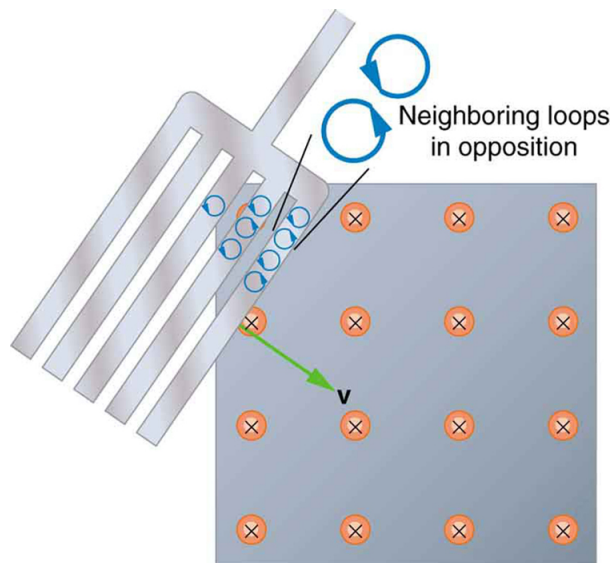
(b) There is little effect on the motion of a slotted metal bob, implying that eddy currents are made less effective. (c) There is also no magnetic damping on a nonconducting bob, since the eddy currents are extremely small.

[\[link\]](#) shows what happens to the metal plate as it enters and leaves the magnetic field. In both cases, it experiences a force opposing its motion. As it enters from the left, flux increases, and so an eddy current is set up (Faraday's law) in the counterclockwise direction (Lenz's law), as shown. Only the right-hand side of the current loop is in the field, so that there is an unopposed force on it to the left (RHR-1). When the metal plate is completely inside the field, there is no eddy current if the field is uniform, since the flux remains constant in this region. But when the plate leaves the field on the right, flux decreases, causing an eddy current in the clockwise direction that, again, experiences a force to the left, further slowing the motion. A similar analysis of what happens when the plate swings from the right toward the left shows that its motion is also damped when entering and leaving the field.



A more detailed look at the conducting plate passing between the poles of a magnet. As it enters and leaves the field, the change in flux produces an eddy current. Magnetic force on the current loop opposes the motion. There is no current and no magnetic drag when the plate is completely inside the uniform field.

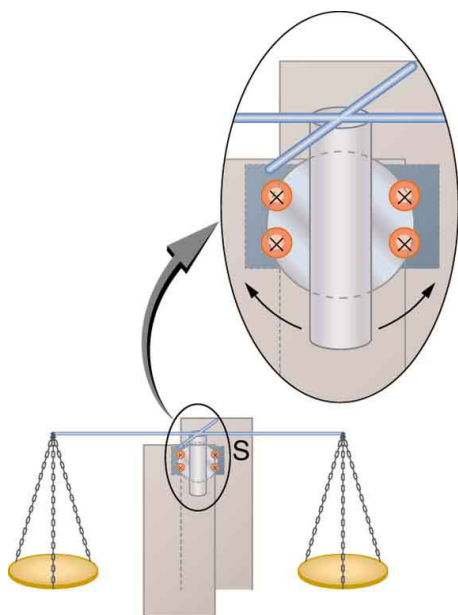
When a slotted metal plate enters the field, as shown in [\[link\]](#), an emf is induced by the change in flux, but it is less effective because the slots limit the size of the current loops. Moreover, adjacent loops have currents in opposite directions, and their effects cancel. When an insulating material is used, the eddy current is extremely small, and so magnetic damping on insulators is negligible. If eddy currents are to be avoided in conductors, then they can be slotted or constructed of thin layers of conducting material separated by insulating sheets.



Eddy currents induced in a slotted metal plate entering a magnetic field form small loops, and the forces on them tend to cancel, thereby making magnetic drag almost zero.

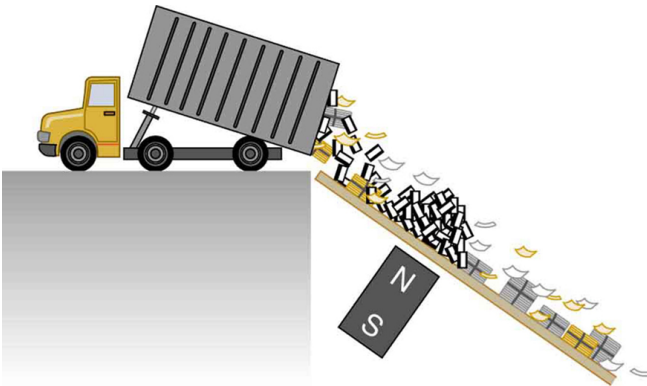
Applications of Magnetic Damping

One use of magnetic damping is found in sensitive laboratory balances. To have maximum sensitivity and accuracy, the balance must be as friction-free as possible. But if it is friction-free, then it will oscillate for a very long time. Magnetic damping is a simple and ideal solution. With magnetic damping, drag is proportional to speed and becomes zero at zero velocity. Thus the oscillations are quickly damped, after which the damping force disappears, allowing the balance to be very sensitive. (See [\[link\]](#).) In most balances, magnetic damping is accomplished with a conducting disc that rotates in a fixed field.



Magnetic damping of this sensitive balance slows its oscillations. Since Faraday's law of induction gives the greatest effect for the most rapid change, damping is greatest for large oscillations and goes to zero as the motion stops.

Since eddy currents and magnetic damping occur only in conductors, recycling centers can use magnets to separate metals from other materials. Trash is dumped in batches down a ramp, beneath which lies a powerful magnet. Conductors in the trash are slowed by magnetic damping while nonmetals in the trash move on, separating from the metals. (See [\[link\]](#).) This works for all metals, not just ferromagnetic ones. A magnet can separate out the ferromagnetic materials alone by acting on stationary trash.



Metals can be separated from other trash by magnetic drag. Eddy currents and magnetic drag are created in the metals sent down this ramp by the powerful magnet beneath it. Nonmetals move on.

Other major applications of eddy currents are in metal detectors and braking systems in trains and roller coasters. Portable metal detectors ([\[link\]](#)) consist of a primary coil carrying an alternating current and a secondary coil in which a current is induced. An eddy current will be induced in a piece of metal close to the detector which will cause a change in the induced current within the secondary coil, leading to some sort of signal like a shrill noise. Braking using eddy currents is safer because factors such as rain do not affect the braking and the braking is smoother. However, eddy currents cannot bring the motion to a complete stop, since the force produced decreases with speed. Thus, speed can be reduced from say 20 m/s to 5 m/s, but another form of braking is needed to completely stop the vehicle. Generally, powerful rare earth magnets such as neodymium magnets are used in roller coasters. [\[link\]](#) shows rows of magnets in such an application. The vehicle has metal fins (normally containing copper) which pass through the magnetic field slowing the vehicle down in much the same way as with the pendulum bob shown in [\[link\]](#).



A soldier in Iraq uses a metal detector to search for explosives and weapons. (credit: U.S. Army)



The rows of rare earth magnets (protruding horizontally) are used for magnetic braking in roller coasters. (credit: Stefan Scheer, Wikimedia Commons)

Induction cooktops have electromagnets under their surface. The magnetic field is varied rapidly producing eddy currents in the base of the pot, causing the pot and its contents to increase in temperature. Induction cooktops have high efficiencies and good response times but the base of the pot needs to be ferromagnetic, iron or steel for induction to work.

Section Summary

- Current loops induced in moving conductors are called eddy currents.
- They can create significant drag, called magnetic damping.

Conceptual Questions

Exercise:

Problem:

Explain why magnetic damping might not be effective on an object made of several thin conducting layers separated by insulation.

Exercise:

Problem:

Explain how electromagnetic induction can be used to detect metals? This technique is particularly important in detecting buried landmines for disposal, geophysical prospecting and at airports.

Problems & Exercises

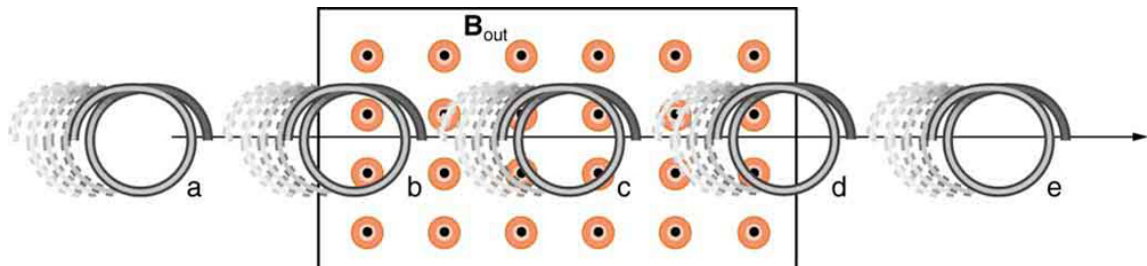
Exercise:

Problem:

Make a drawing similar to [\[link\]](#), but with the pendulum moving in the opposite direction. Then use Faraday's law, Lenz's law, and RHR-1 to show that magnetic force opposes motion.

Exercise:

Problem:



A coil is moved into and out of a region of uniform magnetic field.

A coil is moved through a magnetic field as shown in [\[link\]](#). The field is uniform inside the rectangle and zero outside. What is the direction of the induced current and what is the direction of the magnetic force on the coil at each position shown?

Glossary

eddy current

a current loop in a conductor caused by motional emf

magnetic damping

the drag produced by eddy currents

Electric Generators

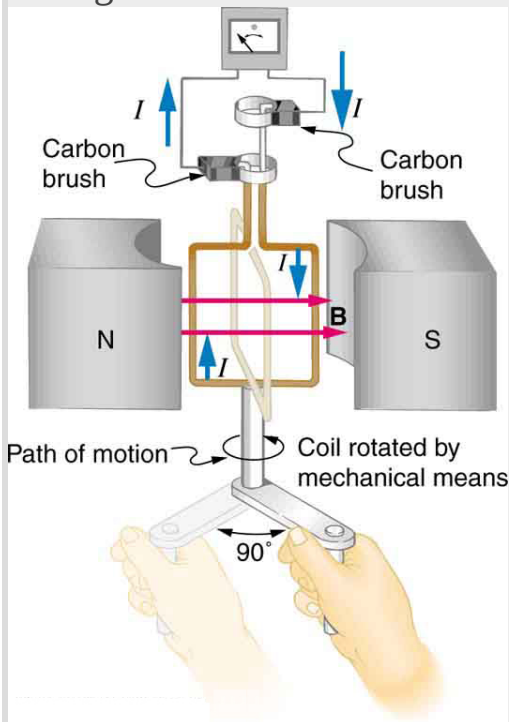
- Calculate the emf induced in a generator.
- Calculate the peak emf which can be induced in a particular generator system.

Electric generators induce an emf by rotating a coil in a magnetic field, as briefly discussed in [Induced Emf and Magnetic Flux](#). We will now explore generators in more detail. Consider the following example.

Example:

Calculating the Emf Induced in a Generator Coil

The generator coil shown in [\[link\]](#) is rotated through one-fourth of a revolution (from $\theta = 0^\circ$ to $\theta = 90^\circ$) in 15.0 ms. The 200-turn circular coil has a 5.00 cm radius and is in a uniform 1.25 T magnetic field. What is the average emf induced?



When this generator coil is rotated through one-fourth of a revolution, the

magnetic flux Φ changes
from its maximum to
zero, inducing an emf.

Strategy

We use Faraday's law of induction to find the average emf induced over a time Δt :

Equation:

$$\text{emf} = -N \frac{\Delta \Phi}{\Delta t}.$$

We know that $N = 200$ and $\Delta t = 15.0$ ms, and so we must determine the change in flux $\Delta \Phi$ to find emf.

Solution

Since the area of the loop and the magnetic field strength are constant, we see that

Equation:

$$\Delta \Phi = \Delta(BA \cos \theta) = AB \Delta(\cos \theta).$$

Now, $\Delta(\cos \theta) = -1.0$, since it was given that θ goes from 0° to 90° .

Thus $\Delta \Phi = -AB$, and

Equation:

$$\text{emf} = N \frac{AB}{\Delta t}.$$

The area of the loop is

$A = \pi r^2 = (3.14...)(0.0500 \text{ m})^2 = 7.85 \times 10^{-3} \text{ m}^2$. Entering this value gives

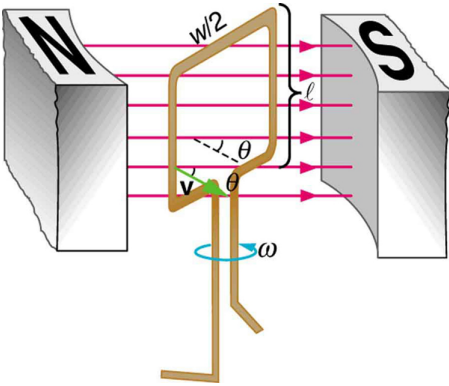
Equation:

$$\text{emf} = 200 \frac{(7.85 \times 10^{-3} \text{ m}^2)(1.25 \text{ T})}{15.0 \times 10^{-3} \text{ s}} = 131 \text{ V}.$$

Discussion

This is a practical average value, similar to the 120 V used in household power.

The emf calculated in [\[link\]](#) is the average over one-fourth of a revolution. What is the emf at any given instant? It varies with the angle between the magnetic field and a perpendicular to the coil. We can get an expression for emf as a function of time by considering the motional emf on a rotating rectangular coil of width w and height ℓ in a uniform magnetic field, as illustrated in [\[link\]](#).



A generator with a single rectangular coil rotated at constant angular velocity in a uniform magnetic field produces an emf that varies sinusoidally in time. Note the generator is similar to a motor, except the shaft is rotated to produce a current rather than the other way around.

Charges in the wires of the loop experience the magnetic force, because they are moving in a magnetic field. Charges in the vertical wires experience forces parallel to the wire, causing currents. But those in the top and bottom segments feel a force perpendicular to the wire, which does not cause a current. We can thus find the induced emf by considering only the side wires. Motional emf is given to be $\text{emf} = B\ell v$, where the velocity v is perpendicular to the magnetic field B . Here the velocity is at an angle θ with B , so that its component perpendicular to B is $v \sin \theta$ (see [\[link\]](#)). Thus in this case the emf induced on each side is $\text{emf} = B\ell v \sin \theta$, and they are in the same direction. The total emf around the loop is then

Equation:

$$\text{emf} = 2B\ell v \sin \theta.$$

This expression is valid, but it does not give emf as a function of time. To find the time dependence of emf, we assume the coil rotates at a constant angular velocity ω . The angle θ is related to angular velocity by $\theta = \omega t$, so that

Equation:

$$\text{emf} = 2B\ell v \sin \omega t.$$

Now, linear velocity v is related to angular velocity ω by $v = r\omega$. Here $r = w/2$, so that $v = (w/2)\omega$, and

Equation:

$$\text{emf} = 2B\ell \frac{w}{2} \omega \sin \omega t = (\ell w) B \omega \sin \omega t.$$

Noting that the area of the loop is $A = \ell w$, and allowing for N loops, we find that

Equation:

$$\text{emf} = NAB\omega \sin \omega t$$

is the **emf induced in a generator coil** of N turns and area A rotating at a constant angular velocity ω in a uniform magnetic field B . This can also be expressed as

Equation:

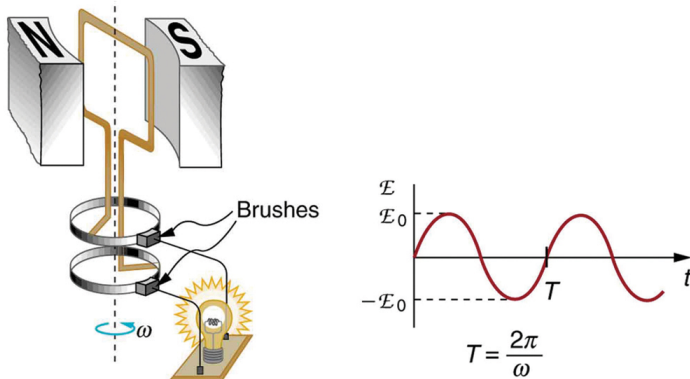
$$\text{emf} = \text{emf}_0 \sin \omega t,$$

where

Equation:

$$\text{emf}_0 = NAB\omega$$

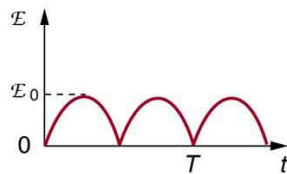
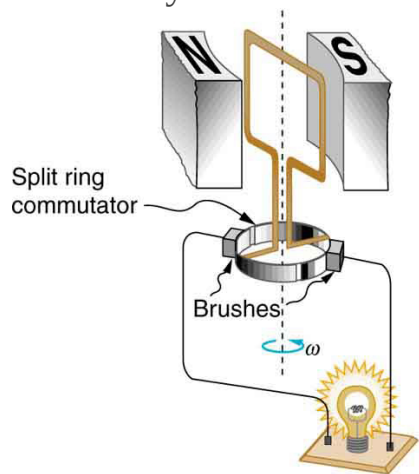
is the maximum **(peak) emf**. Note that the frequency of the oscillation is $f = \omega/2\pi$, and the period is $T = 1/f = 2\pi/\omega$. [\[link\]](#) shows a graph of emf as a function of time, and it now seems reasonable that AC voltage is sinusoidal.



The emf of a generator is sent to a light bulb with the system of rings and brushes shown. The graph gives the emf of the generator as a function of time. emf_0 is the peak emf. The period is $T = 1/f = 2\pi/\omega$, where f is the frequency. Note that the script \mathcal{E} stands for emf.

The fact that the peak emf, $\text{emf}_0 = NAB\omega$, makes good sense. The greater the number of coils, the larger their area, and the stronger the field, the greater the output voltage. It is interesting that the faster the generator is spun (greater ω), the greater the emf. This is noticeable on bicycle generators—at least the cheaper varieties. One of the authors as a juvenile found it amusing to ride his bicycle fast enough to burn out his lights, until he had to ride home lightless one dark night.

[\[link\]](#) shows a scheme by which a generator can be made to produce pulsed DC. More elaborate arrangements of multiple coils and split rings can produce smoother DC, although electronic rather than mechanical means are usually used to make ripple-free DC.



Split rings, called commutators, produce a pulsed DC emf output in this configuration.

Example:**Calculating the Maximum Emf of a Generator**

Calculate the maximum emf, emf_0 , of the generator that was the subject of [\[link\]](#).

Strategy

Once ω , the angular velocity, is determined, $\text{emf}_0 = NAB\omega$ can be used to find emf_0 . All other quantities are known.

Solution

Angular velocity is defined to be the change in angle per unit time:

Equation:

$$\omega = \frac{\Delta\theta}{\Delta t}.$$

One-fourth of a revolution is $\pi/2$ radians, and the time is 0.0150 s; thus,

Equation:

$$\begin{aligned}\omega &= \frac{\pi/2 \text{ rad}}{0.0150 \text{ s}} \\ &= 104.7 \text{ rad/s}.\end{aligned}$$

104.7 rad/s is exactly 1000 rpm. We substitute this value for ω and the information from the previous example into $\text{emf}_0 = NAB\omega$, yielding

Equation:

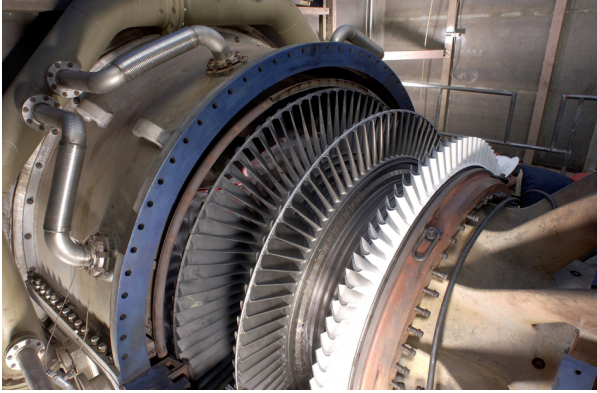
$$\begin{aligned}\text{emf}_0 &= NAB\omega \\ &= 200(7.85 \times 10^{-3} \text{ m}^2)(1.25 \text{ T})(104.7 \text{ rad/s}). \\ &= 206 \text{ V}\end{aligned}$$

Discussion

The maximum emf is greater than the average emf of 131 V found in the previous example, as it should be.

In real life, electric generators look a lot different than the figures in this section, but the principles are the same. The source of mechanical energy that turns the coil can be falling water (hydropower), steam produced by the

burning of fossil fuels, or the kinetic energy of wind. [\[link\]](#) shows a cutaway view of a steam turbine; steam moves over the blades connected to the shaft, which rotates the coil within the generator.



Steam turbine/generator. The steam produced by burning coal impacts the turbine blades, turning the shaft which is connected to the generator.
(credit: Nabonaco, Wikimedia Commons)

Generators illustrated in this section look very much like the motors illustrated previously. This is not coincidental. In fact, a motor becomes a generator when its shaft rotates. Certain early automobiles used their starter motor as a generator. In [Back Emf](#), we shall further explore the action of a motor as a generator.

Section Summary

- An electric generator rotates a coil in a magnetic field, inducing an emf given as a function of time by
Equation:

$$\text{emf} = NAB\omega \sin \omega t,$$

where A is the area of an N -turn coil rotated at a constant angular velocity ω in a uniform magnetic field B .

- The peak emf emf_0 of a generator is

Equation:

$$\text{emf}_0 = NAB\omega.$$

Conceptual Questions

Exercise:

Problem:

Using RHR-1, show that the emfs in the sides of the generator loop in [\[link\]](#) are in the same sense and thus add.

Exercise:

Problem:

The source of a generator's electrical energy output is the work done to turn its coils. How is the work needed to turn the generator related to Lenz's law?

Problems & Exercises

Exercise:

Problem:

Calculate the peak voltage of a generator that rotates its 200-turn, 0.100 m diameter coil at 3600 rpm in a 0.800 T field.

Solution:

474 V

Exercise:

Problem:

At what angular velocity in rpm will the peak voltage of a generator be 480 V, if its 500-turn, 8.00 cm diameter coil rotates in a 0.250 T field?

Exercise:**Problem:**

What is the peak emf generated by rotating a 1000-turn, 20.0 cm diameter coil in the Earth's 5.00×10^{-5} T magnetic field, given the plane of the coil is originally perpendicular to the Earth's field and is rotated to be parallel to the field in 10.0 ms?

Solution:

0.247 V

Exercise:**Problem:**

What is the peak emf generated by a 0.250 m radius, 500-turn coil is rotated one-fourth of a revolution in 4.17 ms, originally having its plane perpendicular to a uniform magnetic field. (This is 60 rev/s.)

Exercise:**Problem:**

(a) A bicycle generator rotates at 1875 rad/s, producing an 18.0 V peak emf. It has a 1.00 by 3.00 cm rectangular coil in a 0.640 T field. How many turns are in the coil? (b) Is this number of turns of wire practical for a 1.00 by 3.00 cm coil?

Solution:

(a) 50

(b) yes

Exercise:**Problem: Integrated Concepts**

This problem refers to the bicycle generator considered in the previous problem. It is driven by a 1.60 cm diameter wheel that rolls on the outside rim of the bicycle tire. (a) What is the velocity of the bicycle if the generator's angular velocity is 1875 rad/s? (b) What is the maximum emf of the generator when the bicycle moves at 10.0 m/s, noting that it was 18.0 V under the original conditions? (c) If the sophisticated generator can vary its own magnetic field, what field strength will it need at 5.00 m/s to produce a 9.00 V maximum emf?

Exercise:**Problem:**

(a) A car generator turns at 400 rpm when the engine is idling. Its 300-turn, 5.00 by 8.00 cm rectangular coil rotates in an adjustable magnetic field so that it can produce sufficient voltage even at low rpms. What is the field strength needed to produce a 24.0 V peak emf? (b) Discuss how this required field strength compares to those available in permanent and electromagnets.

Solution:

(a) 0.477 T

(b) This field strength is small enough that it can be obtained using either a permanent magnet or an electromagnet.

Exercise:**Problem:**

Show that if a coil rotates at an angular velocity ω , the period of its AC output is $2\pi/\omega$.

Exercise:

Problem:

A 75-turn, 10.0 cm diameter coil rotates at an angular velocity of 8.00 rad/s in a 1.25 T field, starting with the plane of the coil parallel to the field. (a) What is the peak emf? (b) At what time is the peak emf first reached? (c) At what time is the emf first at its most negative? (d) What is the period of the AC voltage output?

Solution:

- (a) 5.89 V
- (b) At $t=0$
- (c) 0.393 s
- (d) 0.785 s

Exercise:**Problem:**

(a) If the emf of a coil rotating in a magnetic field is zero at $t = 0$, and increases to its first peak at $t = 0.100$ ms, what is the angular velocity of the coil? (b) At what time will its next maximum occur? (c) What is the period of the output? (d) When is the output first one-fourth of its maximum? (e) When is it next one-fourth of its maximum?

Exercise:**Problem: Unreasonable Results**

A 500-turn coil with a 0.250 m^2 area is spun in the Earth's $5.00 \times 10^{-5} \text{ T}$ field, producing a 12.0 kV maximum emf. (a) At what angular velocity must the coil be spun? (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) $1.92 \times 10^6 \text{ rad/s}$

(b) This angular velocity is unreasonably high, higher than can be obtained for any mechanical system.

(c) The assumption that a voltage as great as 12.0 kV could be obtained is unreasonable.

Glossary

electric generator

a device for converting mechanical work into electric energy; it induces an emf by rotating a coil in a magnetic field

emf induced in a generator coil

$\text{emf} = NAB\omega \sin \omega t$, where A is the area of an N -turn coil rotated at a constant angular velocity ω in a uniform magnetic field B , over a period of time t

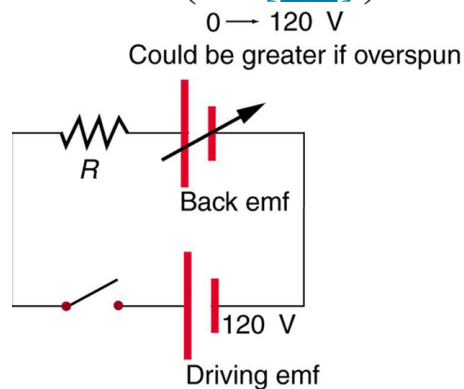
peak emf

$$\text{emf}_0 = NAB\omega$$

Back Emf

- Explain what back emf is and how it is induced.

It has been noted that motors and generators are very similar. Generators convert mechanical energy into electrical energy, whereas motors convert electrical energy into mechanical energy. Furthermore, motors and generators have the same construction. When the coil of a motor is turned, magnetic flux changes, and an emf (consistent with Faraday's law of induction) is induced. The motor thus acts as a generator whenever its coil rotates. This will happen whether the shaft is turned by an external input, like a belt drive, or by the action of the motor itself. That is, when a motor is doing work and its shaft is turning, an emf is generated. Lenz's law tells us the emf opposes any change, so that the input emf that powers the motor will be opposed by the motor's self-generated emf, called the **back emf** of the motor. (See [\[link\]](#).)



The coil of a DC motor is represented as a resistor in this schematic. The back emf is represented as a variable emf that opposes the one driving the motor. Back emf is zero when the motor is not turning, and it increases

proportionally to the
motor's angular
velocity.

Back emf is the generator output of a motor, and so it is proportional to the motor's angular velocity ω . It is zero when the motor is first turned on, meaning that the coil receives the full driving voltage and the motor draws maximum current when it is on but not turning. As the motor turns faster and faster, the back emf grows, always opposing the driving emf, and reduces the voltage across the coil and the amount of current it draws. This effect is noticeable in a number of situations. When a vacuum cleaner, refrigerator, or washing machine is first turned on, lights in the same circuit dim briefly due to the IR drop produced in feeder lines by the large current drawn by the motor. When a motor first comes on, it draws more current than when it runs at its normal operating speed. When a mechanical load is placed on the motor, like an electric wheelchair going up a hill, the motor slows, the back emf drops, more current flows, and more work can be done. If the motor runs at too low a speed, the larger current can overheat it (via resistive power in the coil, $P = I^2 R$), perhaps even burning it out. On the other hand, if there is no mechanical load on the motor, it will increase its angular velocity ω until the back emf is nearly equal to the driving emf. Then the motor uses only enough energy to overcome friction.

Consider, for example, the motor coils represented in [\[link\]](#). The coils have a $0.400\ \Omega$ equivalent resistance and are driven by a $48.0\ \text{V}$ emf. Shortly after being turned on, they draw a current $I = V/R = (48.0\ \text{V}) / (0.400\ \Omega) = 120\ \text{A}$ and, thus, dissipate $P = I^2 R = 5.76\ \text{kW}$ of energy as heat transfer. Under normal operating conditions for this motor, suppose the back emf is $40.0\ \text{V}$. Then at operating speed, the total voltage across the coils is $8.0\ \text{V}$ ($48.0\ \text{V}$ minus the $40.0\ \text{V}$ back emf), and the current drawn is $I = V/R = (8.0\ \text{V}) / (0.400\ \Omega) = 20\ \text{A}$. Under normal load, then, the power dissipated is $P = IV = (20\ \text{A})(8.0\ \text{V}) = 160\ \text{W}$. The latter will not cause a problem for this motor, whereas the former $5.76\ \text{kW}$ would burn out the coils if sustained.

Section Summary

- Any rotating coil will have an induced emf—in motors, this is called back emf, since it opposes the emf input to the motor.

Conceptual Questions

Exercise:

Problem:

Suppose you find that the belt drive connecting a powerful motor to an air conditioning unit is broken and the motor is running freely. Should you be worried that the motor is consuming a great deal of energy for no useful purpose? Explain why or why not.

Problems & Exercises

Exercise:

Problem:

Suppose a motor connected to a 120 V source draws 10.0 A when it first starts. (a) What is its resistance? (b) What current does it draw at its normal operating speed when it develops a 100 V back emf?

Solution:

(a) $12.00\ \Omega$

(b) 1.67 A

Exercise:

Problem:

A motor operating on 240 V electricity has a 180 V back emf at operating speed and draws a 12.0 A current. (a) What is its resistance? (b) What current does it draw when it is first started?

Exercise:**Problem:**

What is the back emf of a 120 V motor that draws 8.00 A at its normal speed and 20.0 A when first starting?

Solution:

72.0 V

Exercise:**Problem:**

The motor in a toy car operates on 6.00 V, developing a 4.50 V back emf at normal speed. If it draws 3.00 A at normal speed, what current does it draw when starting?

Exercise:**Problem: Integrated Concepts**

The motor in a toy car is powered by four batteries in series, which produce a total emf of 6.00 V. The motor draws 3.00 A and develops a 4.50 V back emf at normal speed. Each battery has a $0.100\ \Omega$ internal resistance. What is the resistance of the motor?

Solution:

$0.100\ \Omega$

Glossary**back emf**

the emf generated by a running motor, because it consists of a coil turning in a magnetic field; it opposes the voltage powering the motor

Transformers

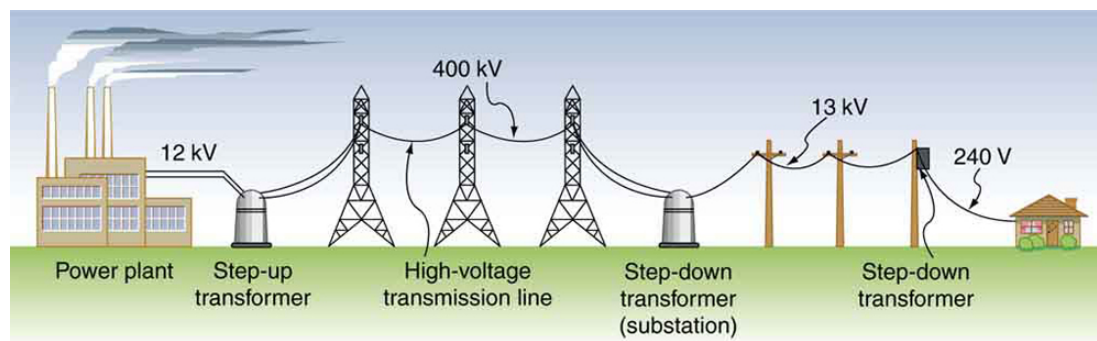
- Explain how a transformer works.
- Calculate voltage, current, and/or number of turns given the other quantities.

Transformers do what their name implies—they transform voltages from one value to another (The term voltage is used rather than emf, because transformers have internal resistance). For example, many cell phones, laptops, video games, and power tools and small appliances have a transformer built into their plug-in unit (like that in [\[link\]](#)) that changes 120 V or 240 V AC into whatever voltage the device uses. Transformers are also used at several points in the power distribution systems, such as illustrated in [\[link\]](#). Power is sent long distances at high voltages, because less current is required for a given amount of power, and this means less line loss, as was discussed previously. But high voltages pose greater hazards, so that transformers are employed to produce lower voltage at the user's location.



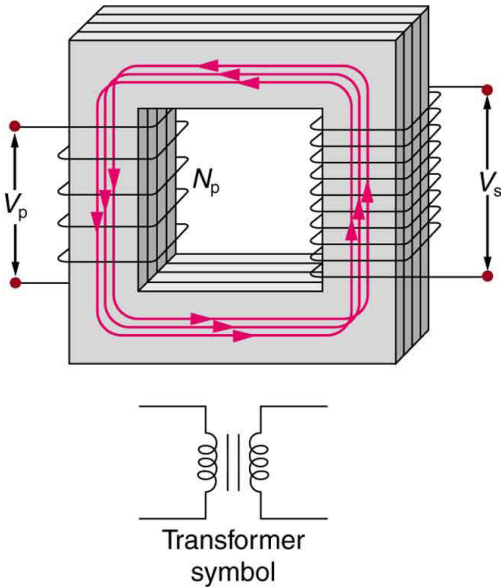
The plug-in transformer has become increasingly familiar with the proliferation of electronic devices that operate on voltages other than common 120 V

AC. Most are in the
3 to 12 V range.
(credit: Shop
Xtreme)



Transformers change voltages at several points in a power distribution system. Electric power is usually generated at greater than 10 kV, and transmitted long distances at voltages over 200 kV—sometimes as great as 700 kV—to limit energy losses. Local power distribution to neighborhoods or industries goes through a substation and is sent short distances at voltages ranging from 5 to 13 kV. This is reduced to 120, 240, or 480 V for safety at the individual user site.

The type of transformer considered in this text—see [\[link\]](#)—is based on Faraday’s law of induction and is very similar in construction to the apparatus Faraday used to demonstrate magnetic fields could cause currents. The two coils are called the *primary* and *secondary coils*. In normal use, the input voltage is placed on the primary, and the secondary produces the transformed output voltage. Not only does the iron core trap the magnetic field created by the primary coil, its magnetization increases the field strength. Since the input voltage is AC, a time-varying magnetic flux is sent to the secondary, inducing its AC output voltage.



A typical construction of a simple transformer has two coils wound on a ferromagnetic core that is laminated to minimize eddy currents. The magnetic field created by the primary is mostly confined to and increased by the core, which transmits it to the secondary coil. Any change in current in the primary induces a current in the secondary.

For the simple transformer shown in [\[link\]](#), the output voltage V_s depends almost entirely on the input voltage V_p and the ratio of the number of loops in the primary and secondary coils. Faraday's law of induction for the secondary coil gives its induced output voltage V_s to be

Equation:

$$V_s = -N_s \frac{\Delta\Phi}{\Delta t},$$

where N_s is the number of loops in the secondary coil and $\Delta\Phi/\Delta t$ is the rate of change of magnetic flux. Note that the output voltage equals the induced emf ($V_s = \text{emf}_s$), provided coil resistance is small (a reasonable assumption for transformers). The cross-sectional area of the coils is the same on either side, as is the magnetic field strength, and so $\Delta\Phi/\Delta t$ is the same on either side. The input primary voltage V_p is also related to changing flux by

Equation:

$$V_p = -N_p \frac{\Delta\Phi}{\Delta t}.$$

The reason for this is a little more subtle. Lenz's law tells us that the primary coil opposes the change in flux caused by the input voltage V_p , hence the minus sign (This is an example of *self-inductance*, a topic to be explored in some detail in later sections). Assuming negligible coil resistance, Kirchhoff's loop rule tells us that the induced emf exactly equals the input voltage. Taking the ratio of these last two equations yields a useful relationship:

Equation:

$$\frac{V_s}{V_p} = \frac{N_s}{N_p}.$$

This is known as the **transformer equation**, and it simply states that the ratio of the secondary to primary voltages in a transformer equals the ratio of the number of loops in their coils.

The output voltage of a transformer can be less than, greater than, or equal to the input voltage, depending on the ratio of the number of loops in their coils. Some transformers even provide a variable output by allowing connection to be made at different points on the secondary coil. A **step-up**

transformer is one that increases voltage, whereas a **step-down transformer** decreases voltage. Assuming, as we have, that resistance is negligible, the electrical power output of a transformer equals its input. This is nearly true in practice—transformer efficiency often exceeds 99%.

Equating the power input and output,

Equation:

$$P_p = I_p V_p = I_s V_s = P_s.$$

Rearranging terms gives

Equation:

$$\frac{V_s}{V_p} = \frac{I_p}{I_s}.$$

Combining this with $\frac{V_s}{V_p} = \frac{N_s}{N_p}$, we find that

Equation:

$$\frac{I_s}{I_p} = \frac{N_p}{N_s}$$

is the relationship between the output and input currents of a transformer. So if voltage increases, current decreases. Conversely, if voltage decreases, current increases.

Example:

Calculating Characteristics of a Step-Up Transformer

A portable x-ray unit has a step-up transformer, the 120 V input of which is transformed to the 100 kV output needed by the x-ray tube. The primary has 50 loops and draws a current of 10.00 A when in use. (a) What is the number of loops in the secondary? (b) Find the current output of the secondary.

Strategy and Solution for (a)

We solve $\frac{V_s}{V_p} = \frac{N_s}{N_p}$ for N_s , the number of loops in the secondary, and enter the known values. This gives

Equation:

$$\begin{aligned} N_s &= N_p \frac{V_s}{V_p} \\ &= (50) \frac{100,000 \text{ V}}{120 \text{ V}} = 4.17 \times 10^4. \end{aligned}$$

Discussion for (a)

A large number of loops in the secondary (compared with the primary) is required to produce such a large voltage. This would be true for neon sign transformers and those supplying high voltage inside TVs and CRTs.

Strategy and Solution for (b)

We can similarly find the output current of the secondary by solving $\frac{I_s}{I_p} = \frac{N_p}{N_s}$ for I_s and entering known values. This gives

Equation:

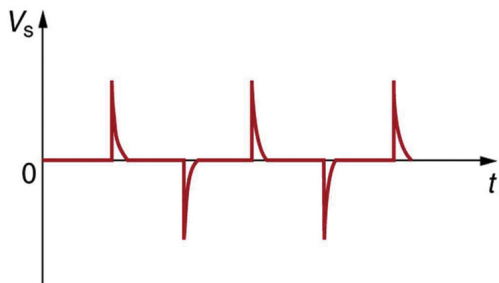
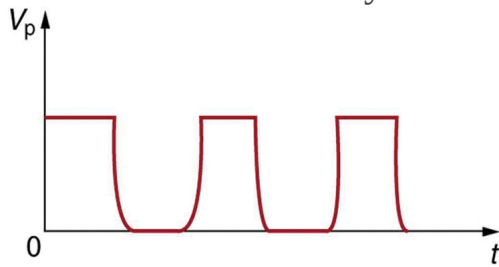
$$\begin{aligned} I_s &= I_p \frac{N_p}{N_s} \\ &= (10.00 \text{ A}) \frac{50}{4.17 \times 10^4} = 12.0 \text{ mA}. \end{aligned}$$

Discussion for (b)

As expected, the current output is significantly less than the input. In certain spectacular demonstrations, very large voltages are used to produce long arcs, but they are relatively safe because the transformer output does not supply a large current. Note that the power input here is $P_p = I_p V_p = (10.00 \text{ A})(120 \text{ V}) = 1.20 \text{ kW}$. This equals the power output $P_p = I_s V_s = (12.0 \text{ mA})(100 \text{ kV}) = 1.20 \text{ kW}$, as we assumed in the derivation of the equations used.

The fact that transformers are based on Faraday's law of induction makes it clear why we cannot use transformers to change DC voltages. If there is no change in primary voltage, there is no voltage induced in the secondary. One possibility is to connect DC to the primary coil through a switch. As the switch is opened and closed, the secondary produces a voltage like that

in [\[link\]](#). This is not really a practical alternative, and AC is in common use wherever it is necessary to increase or decrease voltages.



Transformers do not work for pure DC voltage input, but if it is switched on and off as on the top graph, the output will look something like that on the bottom graph. This is not the sinusoidal AC most AC appliances need.

Example:

Calculating Characteristics of a Step-Down Transformer

A battery charger meant for a series connection of ten nickel-cadmium batteries (total emf of 12.5 V DC) needs to have a 15.0 V output to charge the batteries. It uses a step-down transformer with a 200-loop primary and a 120 V input. (a) How many loops should there be in the secondary coil? (b) If the charging current is 16.0 A, what is the input current?

Strategy and Solution for (a)

You would expect the secondary to have a small number of loops. Solving $\frac{V_s}{V_p} = \frac{N_s}{N_p}$ for N_s and entering known values gives

Equation:

$$\begin{aligned} N_s &= N_p \frac{V_s}{V_p} \\ &= (200) \frac{15.0 \text{ V}}{120 \text{ V}} = 25. \end{aligned}$$

Strategy and Solution for (b)

The current input can be obtained by solving $\frac{I_s}{I_p} = \frac{N_p}{N_s}$ for I_p and entering known values. This gives

Equation:

$$\begin{aligned} I_p &= I_s \frac{N_s}{N_p} \\ &= (16.0 \text{ A}) \frac{25}{200} = 2.00 \text{ A}. \end{aligned}$$

Discussion

The number of loops in the secondary is small, as expected for a step-down transformer. We also see that a small input current produces a larger output current in a step-down transformer. When transformers are used to operate large magnets, they sometimes have a small number of very heavy loops in the secondary. This allows the secondary to have low internal resistance and produce large currents. Note again that this solution is based on the assumption of 100% efficiency—or power out equals power in ($P_p = P_s$)—reasonable for good transformers. In this case the primary and secondary power is 240 W. (Verify this for yourself as a consistency check.) Note that the Ni-Cd batteries need to be charged from a DC power source (as would a 12 V battery). So the AC output of the secondary coil needs to be converted into DC. This is done using something called a rectifier, which uses devices called diodes that allow only a one-way flow of current.

Transformers have many applications in electrical safety systems, which are discussed in [Electrical Safety: Systems and Devices](#).

Note:**PhET Explorations: Generator**

Generate electricity with a bar magnet! Discover the physics behind the phenomena by exploring magnets and how you can use them to make a bulb light.

<https://archive.cnx.org/specials/1e9b7292-ae74-11e5-a9dc-c7c8521ba8e6/generator/#sim-generator>

Section Summary

- Transformers use induction to transform voltages from one value to another.
- For a transformer, the voltages across the primary and secondary coils are related by

Equation:

$$\frac{V_s}{V_p} = \frac{N_s}{N_p},$$

where V_p and V_s are the voltages across primary and secondary coils having N_p and N_s turns.

- The currents I_p and I_s in the primary and secondary coils are related by $\frac{I_s}{I_p} = \frac{N_p}{N_s}$.
- A step-up transformer increases voltage and decreases current, whereas a step-down transformer decreases voltage and increases current.

Conceptual Questions

Exercise:**Problem:**

Explain what causes physical vibrations in transformers at twice the frequency of the AC power involved.

Problems & Exercises

Exercise:

Problem:

A plug-in transformer, like that in [\[link\]](#), supplies 9.00 V to a video game system. (a) How many turns are in its secondary coil, if its input voltage is 120 V and the primary coil has 400 turns? (b) What is its input current when its output is 1.30 A?

Solution:

(a) 30.0

(b) 9.75×10^{-2} A

Exercise:

Problem:

An American traveler in New Zealand carries a transformer to convert New Zealand's standard 240 V to 120 V so that she can use some small appliances on her trip. (a) What is the ratio of turns in the primary and secondary coils of her transformer? (b) What is the ratio of input to output current? (c) How could a New Zealander traveling in the United States use this same transformer to power her 240 V appliances from 120 V?

Exercise:

Problem:

A cassette recorder uses a plug-in transformer to convert 120 V to 12.0 V, with a maximum current output of 200 mA. (a) What is the current input? (b) What is the power input? (c) Is this amount of power reasonable for a small appliance?

Solution:

(a) 20.0 mA

(b) 2.40 W

(c) Yes, this amount of power is quite reasonable for a small appliance.

Exercise:

Problem:

(a) What is the voltage output of a transformer used for rechargeable flashlight batteries, if its primary has 500 turns, its secondary 4 turns, and the input voltage is 120 V? (b) What input current is required to produce a 4.00 A output? (c) What is the power input?

Exercise:

Problem:

(a) The plug-in transformer for a laptop computer puts out 7.50 V and can supply a maximum current of 2.00 A. What is the maximum input current if the input voltage is 240 V? Assume 100% efficiency. (b) If the actual efficiency is less than 100%, would the input current need to be greater or smaller? Explain.

Solution:

(a) 0.063 A

(b) Greater input current needed.

Exercise:

Problem:

A multipurpose transformer has a secondary coil with several points at which a voltage can be extracted, giving outputs of 5.60, 12.0, and 480 V. (a) The input voltage is 240 V to a primary coil of 280 turns. What are the numbers of turns in the parts of the secondary used to produce the output voltages? (b) If the maximum input current is 5.00 A, what are the maximum output currents (each used alone)?

Exercise:**Problem:**

A large power plant generates electricity at 12.0 kV. Its old transformer once converted the voltage to 335 kV. The secondary of this transformer is being replaced so that its output can be 750 kV for more efficient cross-country transmission on upgraded transmission lines.

(a) What is the ratio of turns in the new secondary compared with the old secondary? (b) What is the ratio of new current output to old output (at 335 kV) for the same power? (c) If the upgraded transmission lines have the same resistance, what is the ratio of new line power loss to old?

Solution:

(a) 2.2

(b) 0.45

(c) 0.20, or 20.0%

Exercise:**Problem:**

If the power output in the previous problem is 1000 MW and line resistance is $2.00\ \Omega$, what were the old and new line losses?

Exercise:**Problem: Unreasonable Results**

The 335 kV AC electricity from a power transmission line is fed into the primary coil of a transformer. The ratio of the number of turns in the secondary to the number in the primary is $N_s/N_p = 1000$. (a) What voltage is induced in the secondary? (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) 335 MV

(b) way too high, well beyond the breakdown voltage of air over reasonable distances

(c) input voltage is too high

Exercise:

Problem: Construct Your Own Problem

Consider a double transformer to be used to create very large voltages. The device consists of two stages. The first is a transformer that produces a much larger output voltage than its input. The output of the first transformer is used as input to a second transformer that further increases the voltage. Construct a problem in which you calculate the output voltage of the final stage based on the input voltage of the first stage and the number of turns or loops in both parts of both transformers (four coils in all). Also calculate the maximum output current of the final stage based on the input current. Discuss the possibility of power losses in the devices and the effect on the output current and power.

Glossary

transformer

a device that transforms voltages from one value to another using induction

transformer equation

the equation showing that the ratio of the secondary to primary voltages in a transformer equals the ratio of the number of loops in their coils; $\frac{V_s}{V_p} = \frac{N_s}{N_p}$

step-up transformer

a transformer that increases voltage

step-down transformer
a transformer that decreases voltage

Electrical Safety: Systems and Devices

- Explain how various modern safety features in electric circuits work, with an emphasis on how induction is employed.

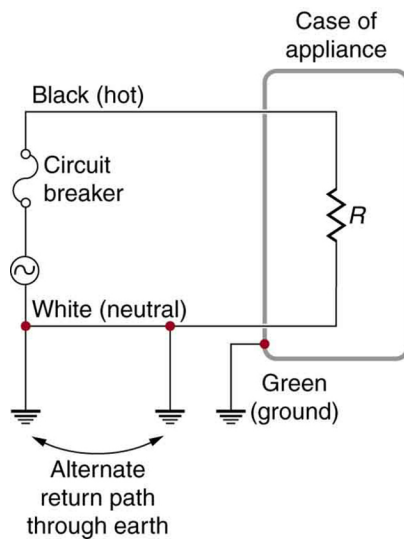
Electricity has two hazards. A **thermal hazard** occurs when there is electrical overheating. A **shock hazard** occurs when electric current passes through a person. Both hazards have already been discussed. Here we will concentrate on systems and devices that prevent electrical hazards.

[\[link\]](#) shows the schematic for a simple AC circuit with no safety features. This is not how power is distributed in practice. Modern household and industrial wiring requires the **three-wire system**, shown schematically in [\[link\]](#), which has several safety features. First is the familiar *circuit breaker* (or *fuse*) to prevent thermal overload. Second, there is a protective *case* around the appliance, such as a toaster or refrigerator. The case's safety feature is that it prevents a person from touching exposed wires and coming into electrical contact with the circuit, helping prevent shocks.



Schematic of a simple AC circuit with a voltage source and a single appliance represented by the resistance R .

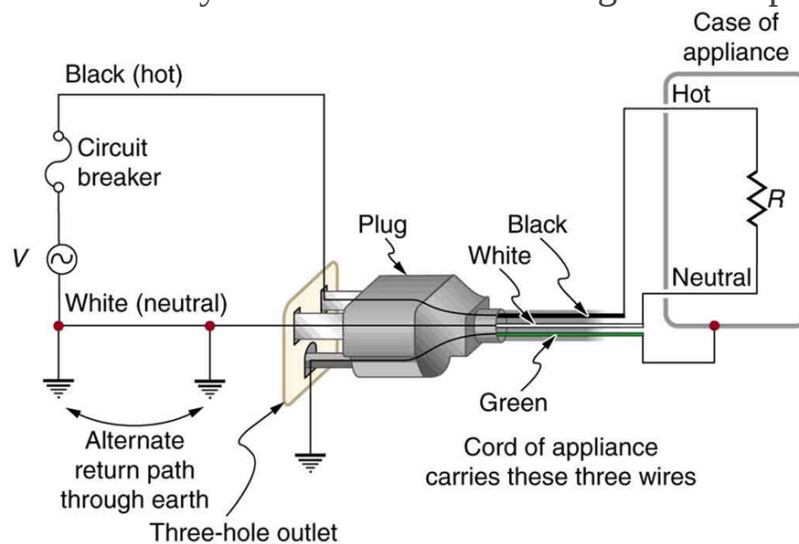
There are no safety features in this circuit.



The three-wire system connects the neutral wire to the earth at the voltage source and user location, forcing it to be at zero volts and supplying an alternative return path for the current through the earth. Also grounded to zero volts is the case of the appliance. A circuit breaker or fuse protects against thermal overload and is in series on the active (live/hot) wire. Note that wire insulation colors vary with region and it is essential to

check locally to determine which color codes are in use (and even if they were followed in the particular installation).

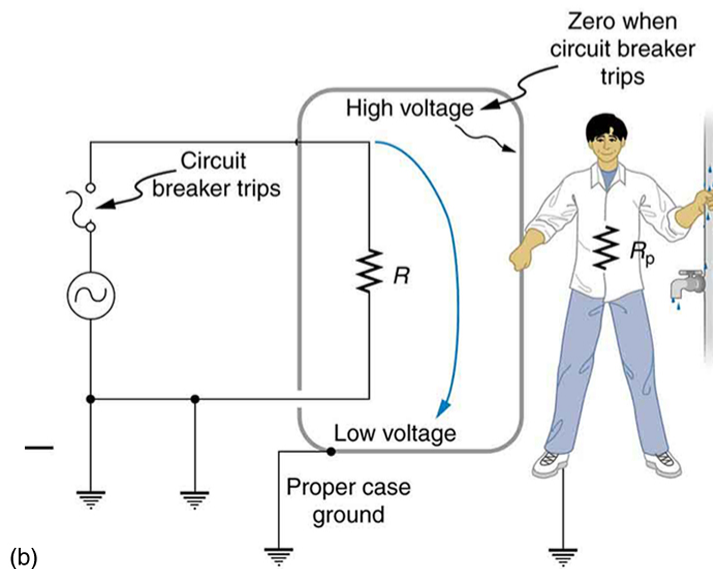
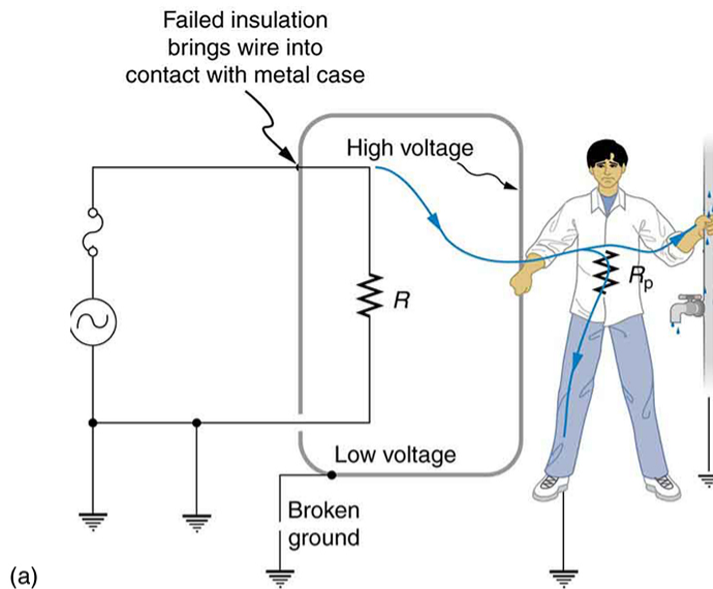
There are *three connections to earth or ground* (hereafter referred to as “earth/ground”) shown in [\[link\]](#). Recall that an earth/ground connection is a low-resistance path directly to the earth. The two earth/ground connections on the *neutral wire* force it to be at zero volts relative to the earth, giving the wire its name. This wire is therefore safe to touch even if its insulation, usually white, is missing. The neutral wire is the return path for the current to follow to complete the circuit. Furthermore, the two earth/ground connections supply an alternative path through the earth, a good conductor, to complete the circuit. The earth/ground connection closest to the power source could be at the generating plant, while the other is at the user’s location. The third earth/ground is to the case of the appliance, through the green *earth/ground wire*, forcing the case, too, to be at zero volts. The *live* or *hot wire* (hereafter referred to as “live/hot”) supplies voltage and current to operate the appliance. [\[link\]](#) shows a more pictorial version of how the three-wire system is connected through a three-prong plug to an appliance.



The standard three-prong plug can only be inserted in one way, to assure proper function of the three-wire system.

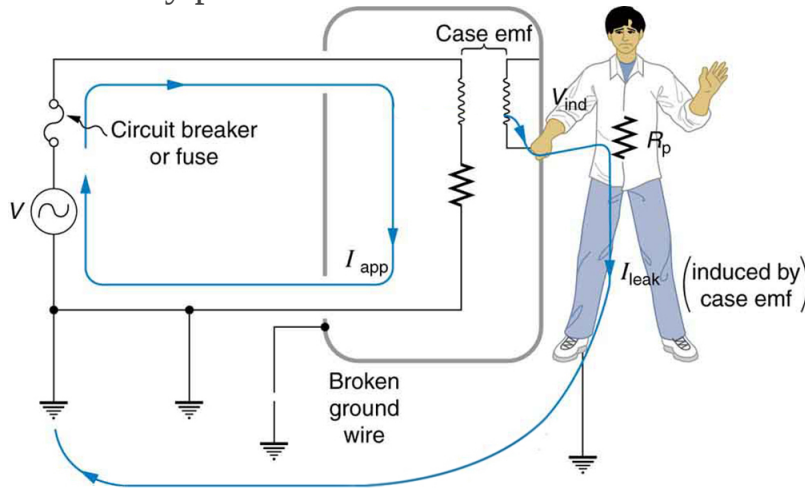
A note on insulation color-coding: Insulating plastic is color-coded to identify live/hot, neutral and ground wires but these codes vary around the world. Live/hot wires may be brown, red, black, blue or grey. Neutral wire may be blue, black or white. Since the same color may be used for live/hot or neutral in different parts of the world, it is essential to determine the color code in your region. The only exception is the earth/ground wire which is often green but may be yellow or just bare wire. Striped coatings are sometimes used for the benefit of those who are colorblind.

The three-wire system replaced the older two-wire system, which lacks an earth/ground wire. Under ordinary circumstances, insulation on the live/hot and neutral wires prevents the case from being directly in the circuit, so that the earth/ground wire may seem like double protection. Grounding the case solves more than one problem, however. The simplest problem is worn insulation on the live/hot wire that allows it to contact the case, as shown in [\[link\]](#). Lacking an earth/ground connection (some people cut the third prong off the plug because they only have outdated two hole receptacles), a severe shock is possible. This is particularly dangerous in the kitchen, where a good connection to earth/ground is available through water on the floor or a water faucet. With the earth/ground connection intact, the circuit breaker will trip, forcing repair of the appliance. Why are some appliances still sold with two-prong plugs? These have nonconducting cases, such as power tools with impact resistant plastic cases, and are called *doubly insulated*. Modern two-prong plugs can be inserted into the asymmetric standard outlet in only one way, to ensure proper connection of live/hot and neutral wires.



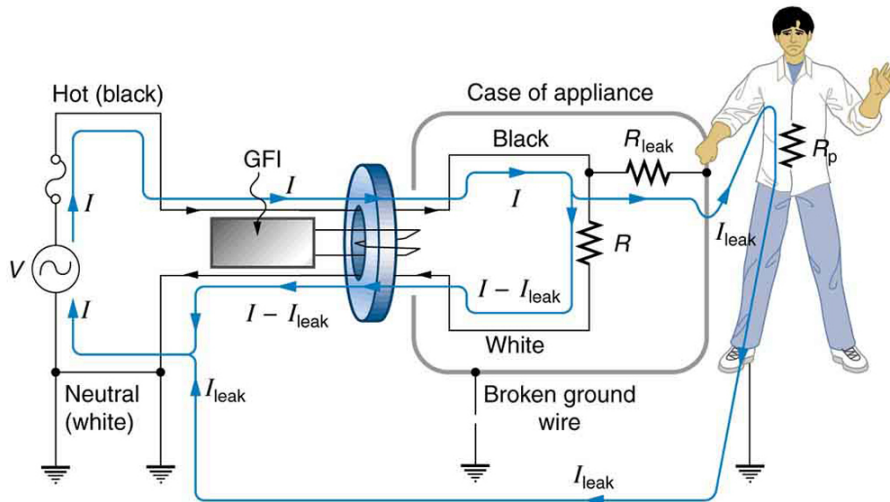
Worn insulation allows the live/hot wire to come into direct contact with the metal case of this appliance. (a) The earth/ground connection being broken, the person is severely shocked. The appliance may operate normally in this situation. (b) With a proper earth/ground, the circuit breaker trips, forcing repair of the appliance.

Electromagnetic induction causes a more subtle problem that is solved by grounding the case. The AC current in appliances can induce an emf on the case. If grounded, the case voltage is kept near zero, but if the case is not grounded, a shock can occur as pictured in [\[link\]](#). Current driven by the induced case emf is called a *leakage current*, although current does not necessarily pass from the resistor to the case.



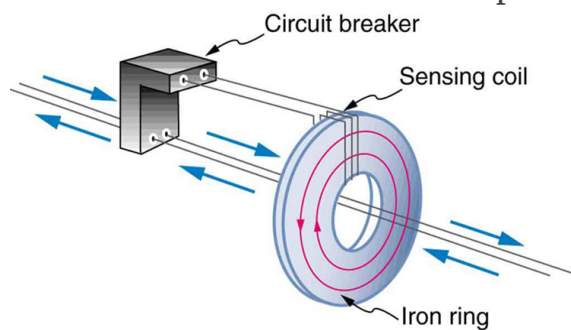
AC currents can induce an emf on the case of an appliance. The voltage can be large enough to cause a shock. If the case is grounded, the induced emf is kept near zero.

A *ground fault interrupter* (GFI) is a safety device found in updated kitchen and bathroom wiring that works based on electromagnetic induction. GFIs compare the currents in the live/hot and neutral wires. When live/hot and neutral currents are not equal, it is almost always because current in the neutral is less than in the live/hot wire. Then some of the current, again called a leakage current, is returning to the voltage source by a path other than through the neutral wire. It is assumed that this path presents a hazard, such as shown in [\[link\]](#). GFIs are usually set to interrupt the circuit if the leakage current is greater than 5 mA, the accepted maximum harmless shock. Even if the leakage current goes safely to earth/ground through an intact earth/ground wire, the GFI will trip, forcing repair of the leakage.



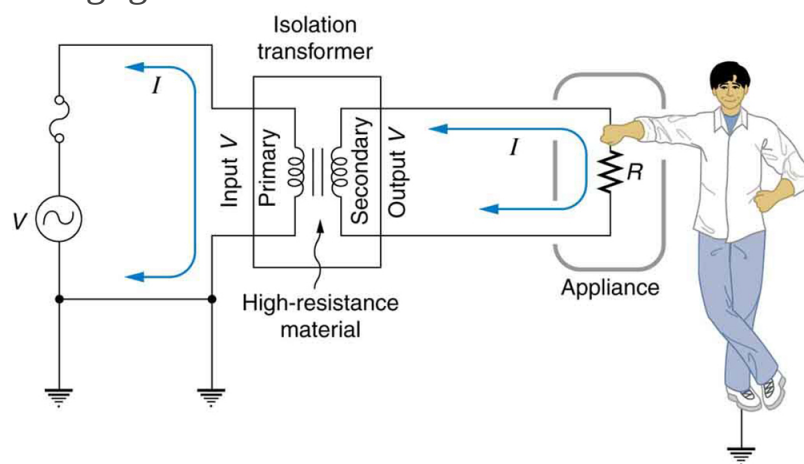
A ground fault interrupter (GFI) compares the currents in the live/hot and neutral wires and will trip if their difference exceeds a safe value. The leakage current here follows a hazardous path that could have been prevented by an intact earth/ground wire.

[\[link\]](#) shows how a GFI works. If the currents in the live/hot and neutral wires are equal, then they induce equal and opposite emfs in the coil. If not, then the circuit breaker will trip.



A GFI compares currents by using both to induce an emf in the same coil. If the currents are equal, they will induce equal but opposite emfs.

Another induction-based safety device is the *isolation transformer*, shown in [\[link\]](#). Most isolation transformers have equal input and output voltages. Their function is to put a large resistance between the original voltage source and the device being operated. This prevents a complete circuit between them, even in the circumstance shown. There is a complete circuit through the appliance. But there is not a complete circuit for current to flow through the person in the figure, who is touching only one of the transformer's output wires, and neither output wire is grounded. The appliance is isolated from the original voltage source by the high resistance of the material between the transformer coils, hence the name isolation transformer. For current to flow through the person, it must pass through the high-resistance material between the coils, through the wire, the person, and back through the earth—a path with such a large resistance that the current is negligible.



An isolation transformer puts a large resistance between the original voltage source and the device, preventing a complete circuit between them.

The basics of electrical safety presented here help prevent many electrical hazards. Electrical safety can be pursued to greater depths. There are, for

example, problems related to different earth/ground connections for appliances in close proximity. Many other examples are found in hospitals. Microshock-sensitive patients, for instance, require special protection. For these people, currents as low as 0.1 mA may cause ventricular fibrillation. The interested reader can use the material presented here as a basis for further study.

Section Summary

- Electrical safety systems and devices are employed to prevent thermal and shock hazards.
- Circuit breakers and fuses interrupt excessive currents to prevent thermal hazards.
- The three-wire system guards against thermal and shock hazards, utilizing live/hot, neutral, and earth/ground wires, and grounding the neutral wire and case of the appliance.
- A ground fault interrupter (GFI) prevents shock by detecting the loss of current to unintentional paths.
- An isolation transformer insulates the device being powered from the original source, also to prevent shock.
- Many of these devices use induction to perform their basic function.

Conceptual Questions

Exercise:

Problem:

Does plastic insulation on live/hot wires prevent shock hazards, thermal hazards, or both?

Exercise:

Problem:

Why are ordinary circuit breakers and fuses ineffective in preventing shocks?

Exercise:

Problem:

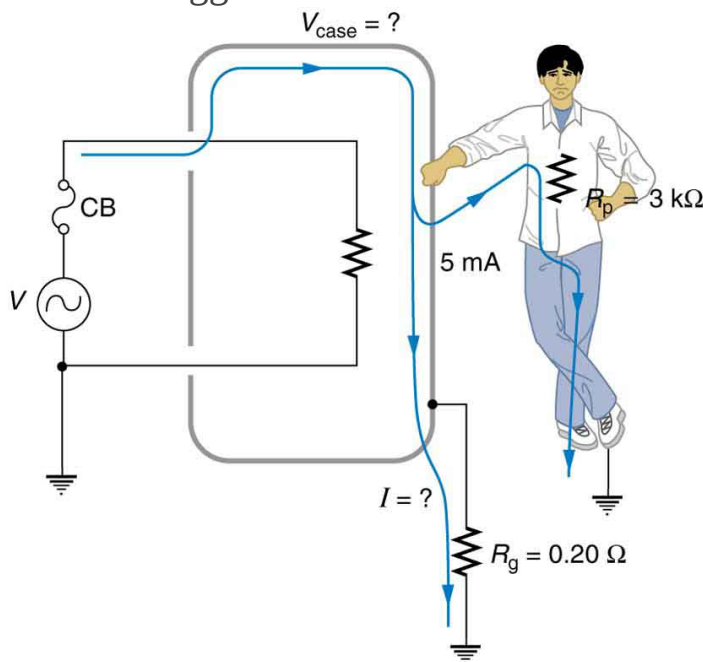
A GFI may trip just because the live/hot and neutral wires connected to it are significantly different in length. Explain why.

Problems & Exercises

Exercise:

Problem: Integrated Concepts

A short circuit to the grounded metal case of an appliance occurs as shown in [\[link\]](#). The person touching the case is wet and only has a $3.00 \text{ k}\Omega$ resistance to earth/ground. (a) What is the voltage on the case if 5.00 mA flows through the person? (b) What is the current in the short circuit if the resistance of the earth/ground wire is 0.200Ω ? (c) Will this trigger the 20.0 A circuit breaker supplying the appliance?



A person can be shocked even when the case of an appliance is grounded.

The large short circuit current

produces a voltage on the case of the appliance, since the resistance of the earth/ground wire is not zero.

Solution:

(a) 15.0 V

(b) 75.0 A

(c) yes

Glossary

thermal hazard

the term for electrical hazards due to overheating

shock hazard

the term for electrical hazards due to current passing through a human

three-wire system

the wiring system used at present for safety reasons, with live, neutral, and ground wires

Introduction to Electromagnetic Waves

class="introduction"

Human eyes
detect these
orange “sea
goldie” fish
swimming
over a coral
reef in the
blue waters
of the Gulf
of Eilat (Red
Sea) using
visible light.

(credit:
Daviddarom
, Wikimedia
Commons)



The beauty of a coral reef, the warm radiance of sunshine, the sting of sunburn, the X-ray revealing a broken bone, even microwave popcorn—all are brought to us by **electromagnetic waves**. The list of the various types of electromagnetic waves, ranging from radio transmission waves to nuclear gamma-ray (γ -ray) emissions, is interesting in itself.

Even more intriguing is that all of these widely varied phenomena are different manifestations of the same thing—electromagnetic waves. (See [\[link\]](#).) What are electromagnetic waves? How are they created, and how do they travel? How can we understand and organize their widely varying properties? What is their relationship to electric and magnetic effects? These and other questions will be explored.

Note:

Misconception Alert: Sound Waves vs. Radio Waves

Many people confuse sound waves with **radio waves**, one type of electromagnetic (EM) wave. However, sound and radio waves are

completely different phenomena. Sound creates pressure variations (waves) in matter, such as air or water, or your eardrum. Conversely, radio waves are *electromagnetic waves*, like visible light, infrared, ultraviolet, X-rays, and gamma rays. EM waves don't need a medium in which to propagate; they can travel through a vacuum, such as outer space. A radio works because sound waves played by the D.J. at the radio station are converted into electromagnetic waves, then encoded and transmitted in the radio-frequency range. The radio in your car receives the radio waves, decodes the information, and uses a speaker to change it back into a sound wave, bringing sweet music to your ears.

Discovering a New Phenomenon

It is worth noting at the outset that the general phenomenon of electromagnetic waves was predicted by theory before it was realized that light is a form of electromagnetic wave. The prediction was made by James Clerk Maxwell in the mid-19th century when he formulated a single theory combining all the electric and magnetic effects known by scientists at that time. "Electromagnetic waves" was the name he gave to the phenomena his theory predicted.

Such a theoretical prediction followed by experimental verification is an indication of the power of science in general, and physics in particular. The underlying connections and unity of physics allow certain great minds to solve puzzles without having all the pieces. The prediction of electromagnetic waves is one of the most spectacular examples of this power. Certain others, such as the prediction of antimatter, will be discussed in later modules.



The
electromagnetic
waves sent
and received by
this 50-foot
radar dish
antenna at
Kennedy Space
Center in
Florida are not
visible, but
help track
expendable
launch vehicles
with high-
definition
imagery. The
first use of this
C-band radar
dish was for
the launch of
the Atlas V
rocket sending
the New
Horizons probe

toward Pluto.
(credit: NASA)

Maxwell's Equations: Electromagnetic Waves Predicted and Observed

- Restate Maxwell's equations.

The Scotsman James Clerk Maxwell (1831–1879) is regarded as the greatest theoretical physicist of the 19th century. (See [\[link\]](#).) Although he died young, Maxwell not only formulated a complete electromagnetic theory, represented by **Maxwell's equations**, he also developed the kinetic theory of gases and made significant contributions to the understanding of color vision and the nature of Saturn's rings.



James Clerk Maxwell, a 19th-century physicist, developed a theory that explained the relationship between electricity and magnetism and correctly predicted that visible light is caused by electromagnetic

waves. (credit:
G. J. Stodart)

Maxwell brought together all the work that had been done by brilliant physicists such as Oersted, Coulomb, Gauss, and Faraday, and added his own insights to develop the overarching theory of electromagnetism. Maxwell's equations are paraphrased here in words because their mathematical statement is beyond the level of this text. However, the equations illustrate how apparently simple mathematical statements can elegantly unite and express a multitude of concepts—why mathematics is the language of science.

Note:

Maxwell's Equations

1. **Electric field lines** originate on positive charges and terminate on negative charges. The electric field is defined as the force per unit charge on a test charge, and the strength of the force is related to the electric constant ϵ_0 , also known as the permittivity of free space. From Maxwell's first equation we obtain a special form of Coulomb's law known as Gauss's law for electricity.
2. **Magnetic field lines** are continuous, having no beginning or end. No magnetic monopoles are known to exist. The strength of the magnetic force is related to the magnetic constant μ_0 , also known as the permeability of free space. This second of Maxwell's equations is known as Gauss's law for magnetism.
3. A changing magnetic field induces an electromotive force (emf) and, hence, an electric field. The direction of the emf opposes the change. This third of Maxwell's equations is Faraday's law of induction, and includes Lenz's law.
4. Magnetic fields are generated by moving charges or by changing electric fields. This fourth of Maxwell's equations encompasses Ampere's law and adds another source of magnetism—changing electric fields.

Maxwell's equations encompass the major laws of electricity and magnetism. What is not so apparent is the symmetry that Maxwell introduced in his mathematical framework. Especially important is his addition of the hypothesis that changing electric fields create magnetic fields. This is exactly analogous (and symmetric) to Faraday's law of induction and had been suspected for some time, but fits beautifully into Maxwell's equations.

Symmetry is apparent in nature in a wide range of situations. In contemporary research, symmetry plays a major part in the search for subatomic particles using massive multinational particle accelerators such as the new Large Hadron Collider at CERN.

Note:**Making Connections: Unification of Forces**

Maxwell's complete and symmetric theory showed that electric and magnetic forces are not separate, but different manifestations of the same thing—the electromagnetic force. This classical unification of forces is one motivation for current attempts to unify the four basic forces in nature—the gravitational, electrical, strong, and weak nuclear forces.

Since changing electric fields create relatively weak magnetic fields, they could not be easily detected at the time of Maxwell's hypothesis. Maxwell realized, however, that oscillating charges, like those in AC circuits, produce changing electric fields. He predicted that these changing fields would propagate from the source like waves generated on a lake by a jumping fish.

The waves predicted by Maxwell would consist of oscillating electric and magnetic fields—defined to be an electromagnetic wave (EM wave). Electromagnetic waves would be capable of exerting forces on charges great distances from their source, and they might thus be detectable. Maxwell calculated that electromagnetic waves would propagate at a speed given by the equation

Equation:

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}.$$

When the values for μ_0 and ϵ_0 are entered into the equation for c , we find that

Equation:

$$c = \frac{1}{\sqrt{(8.85 \times 10^{-12} \frac{\text{C}^2}{\text{N}\cdot\text{m}^2})(4\pi \times 10^{-7} \frac{\text{T}\cdot\text{m}}{\text{A}})}} = 3.00 \times 10^8 \text{ m/s},$$

which is the speed of light. In fact, Maxwell concluded that light is an electromagnetic wave having such wavelengths that it can be detected by the eye.

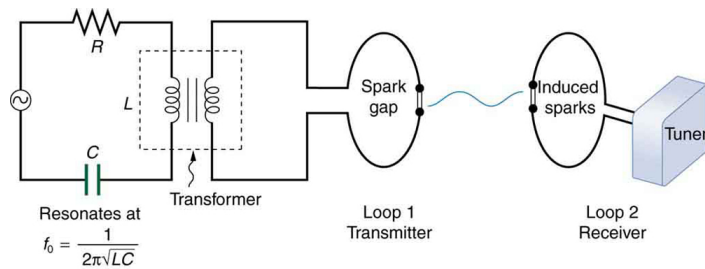
Other wavelengths should exist—it remained to be seen if they did. If so, Maxwell’s theory and remarkable predictions would be verified, the greatest triumph of physics since Newton. Experimental verification came within a few years, but not before Maxwell’s death.

Hertz’s Observations

The German physicist Heinrich Hertz (1857–1894) was the first to generate and detect certain types of electromagnetic waves in the laboratory. Starting in 1887, he performed a series of experiments that not only confirmed the existence of electromagnetic waves, but also verified that they travel at the speed of light.

Hertz used an AC RLC (resistor-inductor-capacitor) circuit that resonates at a known frequency $f_0 = \frac{1}{2\pi\sqrt{LC}}$ and connected it to a loop of wire as shown in [\[link\]](#). High voltages induced across the gap in the loop produced sparks that were visible evidence of the current in the circuit and that helped generate electromagnetic waves.

Across the laboratory, Hertz had another loop attached to another RLC circuit, which could be tuned (as the dial on a radio) to the same resonant frequency as the first and could, thus, be made to receive electromagnetic waves. This loop also had a gap across which sparks were generated, giving solid evidence that electromagnetic waves had been received.



The apparatus used by Hertz in 1887 to generate and detect electromagnetic waves. An RLC circuit connected to the first loop caused sparks across a gap in the wire loop and generated electromagnetic waves. Sparks across a gap in the second loop located across the laboratory gave evidence that the waves had been received.

Hertz also studied the reflection, refraction, and interference patterns of the electromagnetic waves he generated, verifying their wave character. He was able to determine wavelength from the interference patterns, and knowing their frequency, he could calculate the propagation speed using the equation $v = f\lambda$ (velocity—or speed—equals frequency times wavelength). Hertz was thus able to prove that electromagnetic waves travel at the speed of light. The SI unit for frequency, the hertz (1 Hz = 1 cycle/sec), is named in his honor.

Section Summary

- Electromagnetic waves consist of oscillating electric and magnetic fields and propagate at the speed of light c . They were predicted by Maxwell, who also showed that

Equation:

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}},$$

where μ_0 is the permeability of free space and ϵ_0 is the permittivity of free space.

- Maxwell's prediction of electromagnetic waves resulted from his formulation of a complete and symmetric theory of electricity and magnetism, known as Maxwell's equations.
- These four equations are paraphrased in this text, rather than presented numerically, and encompass the major laws of electricity and magnetism. First is Gauss's law for electricity, second is Gauss's law for magnetism, third is Faraday's law of induction, including Lenz's law, and fourth is Ampere's law in a symmetric formulation that adds another source of magnetism—changing electric fields.

Problems & Exercises

Exercise:

Problem:

Verify that the correct value for the speed of light c is obtained when numerical values for the permeability and permittivity of free space (μ_0 and ϵ_0) are entered into the equation $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$.

Exercise:

Problem:

Show that, when SI units for μ_0 and ϵ_0 are entered, the units given by the right-hand side of the equation in the problem above are m/s.

Glossary

electromagnetic waves

radiation in the form of waves of electric and magnetic energy

Maxwell's equations

a set of four equations that comprise a complete, overarching theory of electromagnetism

RLC circuit

an electric circuit that includes a resistor, capacitor and inductor

hertz

an SI unit denoting the frequency of an electromagnetic wave, in cycles per second

speed of light

in a vacuum, such as space, the speed of light is a constant 3×10^8 m/s

electromotive force (emf)

energy produced per unit charge, drawn from a source that produces an electrical current

electric field lines

a pattern of imaginary lines that extend between an electric source and charged objects in the surrounding area, with arrows pointed away from positively charged objects and toward negatively charged objects. The more lines in the pattern, the stronger the electric field in that region

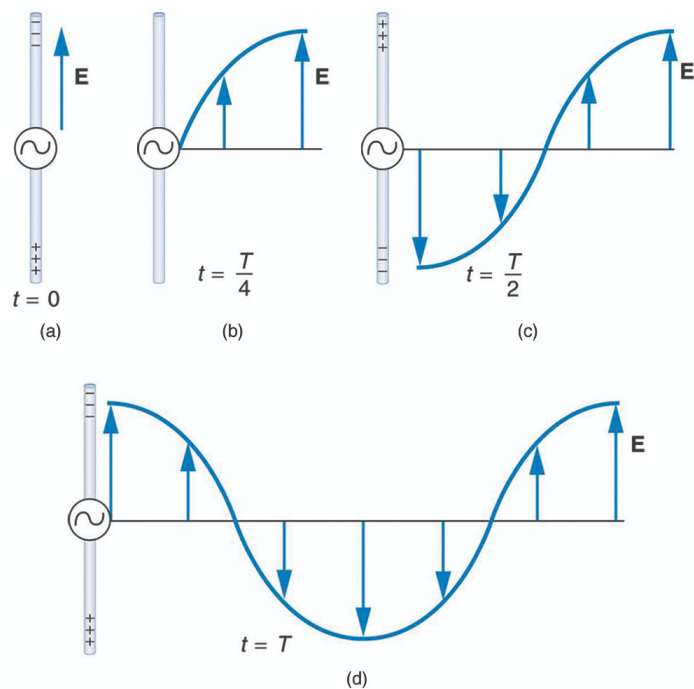
magnetic field lines

a pattern of continuous, imaginary lines that emerge from and enter into opposite magnetic poles. The density of the lines indicates the magnitude of the magnetic field

Production of Electromagnetic Waves

- Describe the electric and magnetic waves as they move out from a source, such as an AC generator.
- Explain the mathematical relationship between the magnetic field strength and the electrical field strength.
- Calculate the maximum strength of the magnetic field in an electromagnetic wave, given the maximum electric field strength.

We can get a good understanding of **electromagnetic waves** (EM) by considering how they are produced. Whenever a current varies, associated electric and magnetic fields vary, moving out from the source like waves. Perhaps the easiest situation to visualize is a varying current in a long straight wire, produced by an AC generator at its center, as illustrated in [\[link\]](#).



This long straight gray wire with an AC generator at its center becomes a broadcast antenna for electromagnetic waves. Shown here are the charge distributions at four different times. The electric field (**E**) propagates away

from the antenna at the speed of light,
forming part of an electromagnetic
wave.

The **electric field** (**E**) shown surrounding the wire is produced by the charge distribution on the wire. Both the **E** and the charge distribution vary as the current changes. The changing field propagates outward at the speed of light.

There is an associated **magnetic field** (**B**) which propagates outward as well (see [\[link\]](#)). The electric and magnetic fields are closely related and propagate as an electromagnetic wave. This is what happens in broadcast antennae such as those in radio and TV stations.

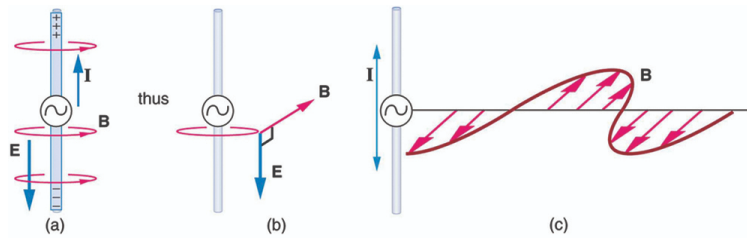
Closer examination of the one complete cycle shown in [\[link\]](#) reveals the periodic nature of the generator-driven charges oscillating up and down in the antenna and the electric field produced. At time $t = 0$, there is the maximum separation of charge, with negative charges at the top and positive charges at the bottom, producing the maximum magnitude of the electric field (or *E*-field) in the upward direction. One-fourth of a cycle later, there is no charge separation and the field next to the antenna is zero, while the maximum *E*-field has moved away at speed c .

As the process continues, the charge separation reverses and the field reaches its maximum downward value, returns to zero, and rises to its maximum upward value at the end of one complete cycle. The outgoing wave has an **amplitude** proportional to the maximum separation of charge. Its **wavelength**(λ) is proportional to the period of the oscillation and, hence, is smaller for short periods or high frequencies. (As usual, wavelength and **frequency**(f) are inversely proportional.)

Electric and Magnetic Waves: Moving Together

Following Ampere's law, current in the antenna produces a magnetic field, as shown in [\[link\]](#). The relationship between **E** and **B** is shown at one

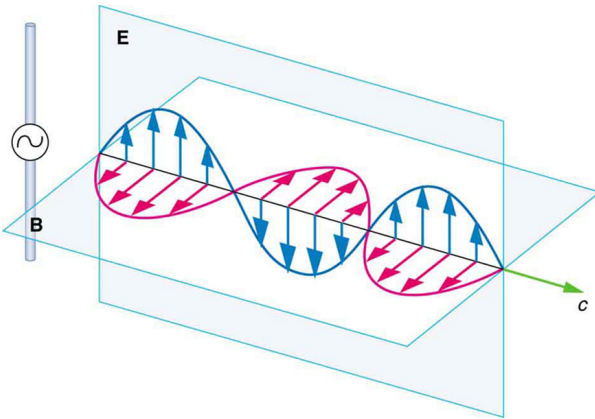
instant in [\[link\]](#) (a). As the current varies, the magnetic field varies in magnitude and direction.



(a) The current in the antenna produces the circular magnetic field lines. The current (I) produces the separation of charge along the wire, which in turn creates the electric field as shown. (b) The electric and magnetic fields (\mathbf{E} and \mathbf{B}) near the wire are perpendicular; they are shown here for one point in space. (c) The magnetic field varies with current and propagates away from the antenna at the speed of light.

The magnetic field lines also propagate away from the antenna at the speed of light, forming the other part of the electromagnetic wave, as seen in [\[link\]](#) (b). The magnetic part of the wave has the same period and wavelength as the electric part, since they are both produced by the same movement and separation of charges in the antenna.

The electric and magnetic waves are shown together at one instant in time in [\[link\]](#). The electric and magnetic fields produced by a long straight wire antenna are exactly in phase. Note that they are perpendicular to one another and to the direction of propagation, making this a **transverse wave**.



A part of the electromagnetic wave sent out from the antenna at one instant in time. The electric and magnetic fields (**E** and **B**) are in phase, and they are perpendicular to one another and the direction of propagation. For clarity, the waves are shown only along one direction, but they propagate out in other directions too.

Electromagnetic waves generally propagate out from a source in all directions, sometimes forming a complex radiation pattern. A linear antenna like this one will not radiate parallel to its length, for example. The wave is shown in one direction from the antenna in [\[link\]](#) to illustrate its basic characteristics.

Instead of the AC generator, the antenna can also be driven by an AC circuit. In fact, charges radiate whenever they are accelerated. But while a current in a circuit needs a complete path, an antenna has a varying charge distribution forming a **standing wave**, driven by the AC. The dimensions of the antenna are critical for determining the frequency of the radiated electromagnetic waves. This is a **resonant** phenomenon and when we tune

radios or TV, we vary electrical properties to achieve appropriate resonant conditions in the antenna.

Receiving Electromagnetic Waves

Electromagnetic waves carry energy away from their source, similar to a sound wave carrying energy away from a standing wave on a guitar string. An antenna for receiving EM signals works in reverse. And like antennas that produce EM waves, receiver antennas are specially designed to resonate at particular frequencies.

An incoming electromagnetic wave accelerates electrons in the antenna, setting up a standing wave. If the radio or TV is switched on, electrical components pick up and amplify the signal formed by the accelerating electrons. The signal is then converted to audio and/or video format. Sometimes big receiver dishes are used to focus the signal onto an antenna.

In fact, charges radiate whenever they are accelerated. When designing circuits, we often assume that energy does not quickly escape AC circuits, and mostly this is true. A broadcast antenna is specially designed to enhance the rate of electromagnetic radiation, and shielding is necessary to keep the radiation close to zero. Some familiar phenomena are based on the production of electromagnetic waves by varying currents. Your microwave oven, for example, sends electromagnetic waves, called microwaves, from a concealed antenna that has an oscillating current imposed on it.

Relating E -Field and B -Field Strengths

There is a relationship between the E - and B -field strengths in an electromagnetic wave. This can be understood by again considering the antenna just described. The stronger the E -field created by a separation of charge, the greater the current and, hence, the greater the B -field created.

Since current is directly proportional to voltage (Ohm's law) and voltage is directly proportional to E -field strength, the two should be directly proportional. It can be shown that the magnitudes of the fields do have a constant ratio, equal to the speed of light. That is,

Equation:

$$\frac{E}{B} = c$$

is the ratio of E -field strength to B -field strength in any electromagnetic wave. This is true at all times and at all locations in space. A simple and elegant result.

Example:**Calculating B -Field Strength in an Electromagnetic Wave**

What is the maximum strength of the B -field in an electromagnetic wave that has a maximum E -field strength of 1000 V/m?

Strategy

To find the B -field strength, we rearrange the above equation to solve for B , yielding

Equation:

$$B = \frac{E}{c}.$$

Solution

We are given E , and c is the speed of light. Entering these into the expression for B yields

Equation:

$$B = \frac{1000 \text{ V/m}}{3.00 \times 10^8 \text{ m/s}} = 3.33 \times 10^{-6} \text{ T},$$

Where T stands for Tesla, a measure of magnetic field strength.

Discussion

The B -field strength is less than a tenth of the Earth's admittedly weak magnetic field. This means that a relatively strong electric field of 1000 V/m is accompanied by a relatively weak magnetic field. Note that as this

wave spreads out, say with distance from an antenna, its field strengths become progressively weaker.

The result of this example is consistent with the statement made in the module [Maxwell's Equations: Electromagnetic Waves Predicted and Observed](#) that changing electric fields create relatively weak magnetic fields. They can be detected in electromagnetic waves, however, by taking advantage of the phenomenon of resonance, as Hertz did. A system with the same natural frequency as the electromagnetic wave can be made to oscillate. All radio and TV receivers use this principle to pick up and then amplify weak electromagnetic waves, while rejecting all others not at their resonant frequency.

Note:

Take-Home Experiment: Antennas

For your TV or radio at home, identify the antenna, and sketch its shape. If you don't have cable, you might have an outdoor or indoor TV antenna. Estimate its size. If the TV signal is between 60 and 216 MHz for basic channels, then what is the wavelength of those EM waves?

Try tuning the radio and note the small range of frequencies at which a reasonable signal for that station is received. (This is easier with digital readout.) If you have a car with a radio and extendable antenna, note the quality of reception as the length of the antenna is changed.

Note:

PhET Explorations: Radio Waves and Electromagnetic Fields

Broadcast radio waves from KPhET. Wiggle the transmitter electron manually or have it oscillate automatically. Display the field as a curve or vectors. The strip chart shows the electron positions at the transmitter and at the receiver.

<https://archive.cnx.org/specials/c8dd764c-ae74-11e5-af4c-3375261fa183/radio-waves/#sim-radio-waves>

Section Summary

- Electromagnetic waves are created by oscillating charges (which radiate whenever accelerated) and have the same frequency as the oscillation.
- Since the electric and magnetic fields in most electromagnetic waves are perpendicular to the direction in which the wave moves, it is ordinarily a transverse wave.
- The strengths of the electric and magnetic parts of the wave are related by
Equation:

$$\frac{E}{B} = c,$$

which implies that the magnetic field B is very weak relative to the electric field E .

Conceptual Questions

Exercise:

Problem:

The direction of the electric field shown in each part of [\[link\]](#) is that produced by the charge distribution in the wire. Justify the direction shown in each part, using the Coulomb force law and the definition of $\mathbf{E} = \mathbf{F}/q$, where q is a positive test charge.

Exercise:

Problem:

Is the direction of the magnetic field shown in [\[link\]](#) (a) consistent with the right-hand rule for current (RHR-2) in the direction shown in the figure?

Exercise:

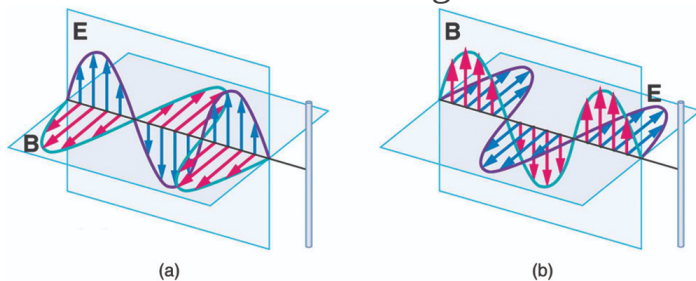
Problem:

Why is the direction of the current shown in each part of [\[link\]](#) opposite to the electric field produced by the wire's charge separation?

Exercise:

Problem:

In which situation shown in [\[link\]](#) will the electromagnetic wave be more successful in inducing a current in the wire? Explain.

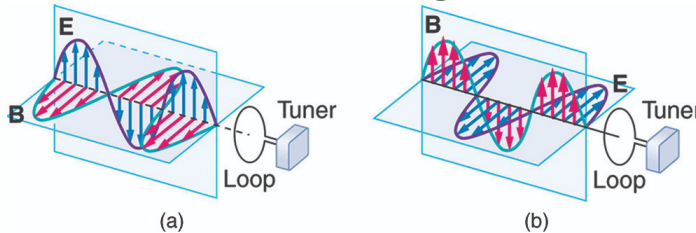


Electromagnetic waves approaching long straight wires.

Exercise:

Problem:

In which situation shown in [\[link\]](#) will the electromagnetic wave be more successful in inducing a current in the loop? Explain.



Electromagnetic waves approaching a wire loop.

Exercise:**Problem:**

Should the straight wire antenna of a radio be vertical or horizontal to best receive radio waves broadcast by a vertical transmitter antenna? How should a loop antenna be aligned to best receive the signals? (Note that the direction of the loop that produces the best reception can be used to determine the location of the source. It is used for that purpose in tracking tagged animals in nature studies, for example.)

Exercise:**Problem:**

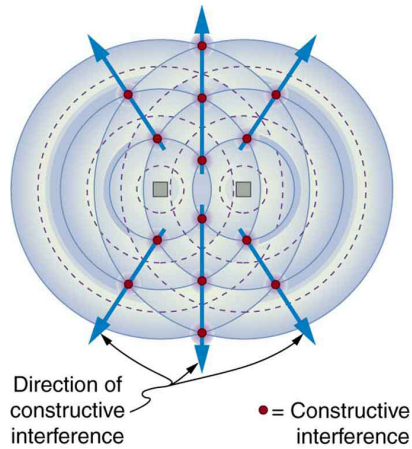
Under what conditions might wires in a DC circuit emit electromagnetic waves?

Exercise:

Problem: Give an example of interference of electromagnetic waves.

Exercise:**Problem:**

[\[link\]](#) shows the interference pattern of two radio antennas broadcasting the same signal. Explain how this is analogous to the interference pattern for sound produced by two speakers. Could this be used to make a directional antenna system that broadcasts preferentially in certain directions? Explain.



An overhead view
of two radio
broadcast antennas
sending the same
signal, and the
interference pattern
they produce.

Exercise:

Problem: Can an antenna be any length? Explain your answer.

Problems & Exercises

Exercise:

Problem:

What is the maximum electric field strength in an electromagnetic wave that has a maximum magnetic field strength of $5.00 \times 10^{-4} \text{ T}$ (about 10 times the Earth's)?

Solution:

150 kV/m

Exercise:

Problem:

The maximum magnetic field strength of an electromagnetic field is 5×10^{-6} T. Calculate the maximum electric field strength if the wave is traveling in a medium in which the speed of the wave is $0.75c$.

Exercise:

Problem:

Verify the units obtained for magnetic field strength B in [\[link\]](#) (using the equation $B = \frac{E}{c}$) are in fact teslas (T).

Glossary

electric field

a vector quantity (**E**); the lines of electric force per unit charge, moving radially outward from a positive charge and in toward a negative charge

electric field strength

the magnitude of the electric field, denoted E -field

magnetic field

a vector quantity (**B**); can be used to determine the magnetic force on a moving charged particle

magnetic field strength

the magnitude of the magnetic field, denoted B -field

transverse wave

a wave, such as an electromagnetic wave, which oscillates perpendicular to the axis along the line of travel

standing wave

a wave that oscillates in place, with nodes where no motion happens

wavelength

the distance from one peak to the next in a wave

amplitude

the height, or magnitude, of an electromagnetic wave

frequency

the number of complete wave cycles (up-down-up) passing a given point within one second (cycles/second)

resonant

a system that displays enhanced oscillation when subjected to a periodic disturbance of the same frequency as its natural frequency

oscillate

to fluctuate back and forth in a steady beat

The Electromagnetic Spectrum

- List three “rules of thumb” that apply to the different frequencies along the electromagnetic spectrum.
- Explain why the higher the frequency, the shorter the wavelength of an electromagnetic wave.
- Draw a simplified electromagnetic spectrum, indicating the relative positions, frequencies, and spacing of the different types of radiation bands.
- List and explain the different methods by which electromagnetic waves are produced across the spectrum.

In this module we examine how electromagnetic waves are classified into categories such as radio, infrared, ultraviolet, and so on, so that we can understand some of their similarities as well as some of their differences. We will also find that there are many connections with previously discussed topics, such as wavelength and resonance. A brief overview of the production and utilization of electromagnetic waves is found in [\[link\]](#).

Type of EM wave	Production	Applications	Life sciences aspect	Issues
Radio & TV	Accelerating charges	Communications Remote controls	MRI	Requires controls for band use
Microwaves	Accelerating charges & thermal agitation	Communications Ovens Radar	Deep heating	Cell phone use
Infrared	Thermal agitations & electronic transitions	Thermal imaging Heating	Absorbed by atmosphere	Greenhouse effect
Visible light	Thermal agitations & electronic transitions	All pervasive	Photosynthesis Human vision	

Type of EM wave	Production	Applications	Life sciences aspect	Issues
Ultraviolet	Thermal agitations & electronic transitions	Sterilization Cancer control	Vitamin D production	Ozone depletion Cancer causing
X-rays	Inner electronic transitions and fast collisions	Medical Security	Medical diagnosis Cancer therapy	Cancer causing
Gamma rays	Nuclear decay	Nuclear medicineSecurity	Medical diagnosis Cancer therapy	Cancer causing Radiation damage

Electromagnetic Waves

Note:

Connections: Waves

There are many types of waves, such as water waves and even earthquakes. Among the many shared attributes of waves are propagation speed, frequency, and wavelength. These are always related by the expression $v_W = f\lambda$. This module concentrates on EM waves, but other modules contain examples of all of these characteristics for sound waves and submicroscopic particles.

As noted before, an electromagnetic wave has a frequency and a wavelength associated with it and travels at the speed of light, or c . The relationship among these wave characteristics can be described by $v_W = f\lambda$, where v_W is the propagation speed of the wave, f is the frequency, and λ is the wavelength. Here $v_W = c$, so that for all electromagnetic waves,

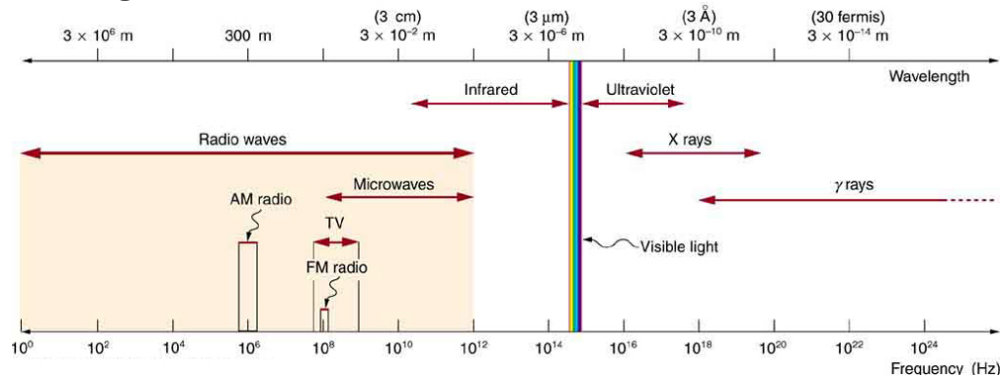
Equation:

$$c = f\lambda.$$

Thus, for all electromagnetic waves, the greater the frequency, the smaller the wavelength.

[\[link\]](#) shows how the various types of electromagnetic waves are categorized according to their wavelengths and frequencies—that is, it shows the electromagnetic spectrum. Many of the

characteristics of the various types of electromagnetic waves are related to their frequencies and wavelengths, as we shall see.



The electromagnetic spectrum, showing the major categories of electromagnetic waves. The range of frequencies and wavelengths is remarkable. The dividing line between some categories is distinct, whereas other categories overlap.

Note:

Electromagnetic Spectrum: Rules of Thumb

Three rules that apply to electromagnetic waves in general are as follows:

- High-frequency electromagnetic waves are more energetic and are more able to penetrate than low-frequency waves.
- High-frequency electromagnetic waves can carry more information per unit time than low-frequency waves.
- The shorter the wavelength of any electromagnetic wave probing a material, the smaller the detail it is possible to resolve.

Note that there are exceptions to these rules of thumb.

Transmission, Reflection, and Absorption

What happens when an electromagnetic wave impinges on a material? If the material is transparent to the particular frequency, then the wave can largely be transmitted. If the material is opaque to the frequency, then the wave can be totally reflected. The wave can also be absorbed by the material, indicating that there is some interaction between the wave and the material, such as the thermal agitation of molecules.

Of course it is possible to have partial transmission, reflection, and absorption. We normally associate these properties with visible light, but they do apply to all electromagnetic waves.

What is not obvious is that something that is transparent to light may be opaque at other frequencies. For example, ordinary glass is transparent to visible light but largely opaque to ultraviolet radiation. Human skin is opaque to visible light—we cannot see through people—but transparent to X-rays.

Radio and TV Waves

The broad category of **radio waves** is defined to contain any electromagnetic wave produced by currents in wires and circuits. Its name derives from their most common use as a carrier of audio information (i.e., radio). The name is applied to electromagnetic waves of similar frequencies regardless of source. Radio waves from outer space, for example, do not come from alien radio stations. They are created by many astronomical phenomena, and their study has revealed much about nature on the largest scales.

There are many uses for radio waves, and so the category is divided into many subcategories, including microwaves and those electromagnetic waves used for AM and FM radio, cellular telephones, and TV.

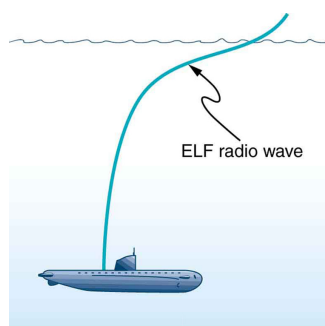
The lowest commonly encountered radio frequencies are produced by high-voltage AC power transmission lines at frequencies of 50 or 60 Hz. (See [\[link\]](#).) These extremely long wavelength electromagnetic waves (about 6000 km!) are one means of energy loss in long-distance power transmission.



This high-voltage traction power line running to Eutingen Railway Substation in Germany radiates electromagnetic waves with very long wavelengths. (credit: Zonk43, Wikimedia Commons)

There is an ongoing controversy regarding potential health hazards associated with exposure to these electromagnetic fields (*E*-fields). Some people suspect that living near such transmission lines may cause a variety of illnesses, including cancer. But demographic data are either inconclusive or simply do not support the hazard theory. Recent reports that have looked at many European and American epidemiological studies have found no increase in risk for cancer due to exposure to *E*-fields.

Extremely low frequency (ELF) radio waves of about 1 kHz are used to communicate with submerged submarines. The ability of radio waves to penetrate salt water is related to their wavelength (much like ultrasound penetrating tissue)—the longer the wavelength, the farther they penetrate. Since salt water is a good conductor, radio waves are strongly absorbed by it, and very long wavelengths are needed to reach a submarine under the surface. (See [\[link\]](#).)



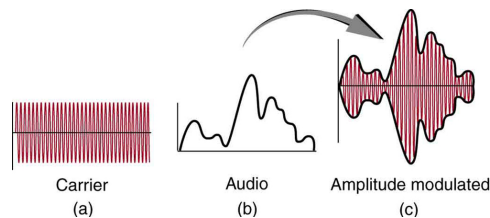
Very long wavelength radio waves are needed to reach this submarine, requiring extremely low frequency signals (ELF).

Shorter wavelengths do not penetrate to any significant depth.

AM radio waves are used to carry commercial radio signals in the frequency range from 540 to 1600 kHz. The abbreviation AM stands for **amplitude modulation**, which is the method for placing information on these waves. (See [\[link\]](#).) A **carrier wave** having the basic frequency of the radio station, say 1530 kHz, is varied or modulated in amplitude by an audio signal. The resulting wave has a constant frequency, but a varying amplitude.

A radio receiver tuned to have the same resonant frequency as the carrier wave can pick up the signal, while rejecting the many other frequencies impinging on its antenna. The receiver's

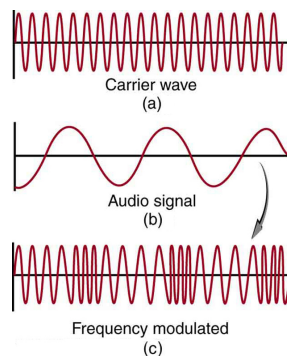
circuitry is designed to respond to variations in amplitude of the carrier wave to replicate the original audio signal. That audio signal is amplified to drive a speaker or perhaps to be recorded.



Amplitude modulation for AM radio. (a) A carrier wave at the station's basic frequency. (b) An audio signal at much lower audible frequencies. (c) The amplitude of the carrier is modulated by the audio signal without changing its basic frequency.

FM Radio Waves

FM radio waves are also used for commercial radio transmission, but in the frequency range of 88 to 108 MHz. FM stands for **frequency modulation**, another method of carrying information. (See [\[link\]](#).) Here a carrier wave having the basic frequency of the radio station, perhaps 105.1 MHz, is modulated in frequency by the audio signal, producing a wave of constant amplitude but varying frequency.



Frequency
modulation for

FM radio. (a) A carrier wave at the station's basic frequency. (b) An audio signal at much lower audible frequencies. (c) The frequency of the carrier is modulated by the audio signal without changing its amplitude.

Since audible frequencies range up to 20 kHz (or 0.020 MHz) at most, the frequency of the FM radio wave can vary from the carrier by as much as 0.020 MHz. Thus the carrier frequencies of two different radio stations cannot be closer than 0.020 MHz. An FM receiver is tuned to resonate at the carrier frequency and has circuitry that responds to variations in frequency, reproducing the audio information.

FM radio is inherently less subject to noise from stray radio sources than AM radio. The reason is that amplitudes of waves add. So an AM receiver would interpret noise added onto the amplitude of its carrier wave as part of the information. An FM receiver can be made to reject amplitudes other than that of the basic carrier wave and only look for variations in frequency. It is thus easier to reject noise from FM, since noise produces a variation in amplitude.

Television is also broadcast on electromagnetic waves. Since the waves must carry a great deal of visual as well as audio information, each channel requires a larger range of frequencies than simple radio transmission. TV channels utilize frequencies in the range of 54 to 88 MHz and 174 to 222 MHz. (The entire FM radio band lies between channels 88 MHz and 174 MHz.) These TV channels are called VHF (for **very high frequency**). Other channels called UHF (for **ultra high frequency**) utilize an even higher frequency range of 470 to 1000 MHz.

The TV video signal is AM, while the TV audio is FM. Note that these frequencies are those of free transmission with the user utilizing an old-fashioned roof antenna. Satellite dishes and cable transmission of TV occurs at significantly higher frequencies and is rapidly evolving with the use of the high-definition or HD format.

Example: **Calculating Wavelengths of Radio Waves**

Calculate the wavelengths of a 1530-kHz AM radio signal, a 105.1-MHz FM radio signal, and a 1.90-GHz cell phone signal.

Strategy

The relationship between wavelength and frequency is $c = f\lambda$, where $c = 3.00 \times 10^8$ m/s is the speed of light (the speed of light is only very slightly smaller in air than it is in a vacuum). We can rearrange this equation to find the wavelength for all three frequencies.

Solution

Rearranging gives

Equation:

$$\lambda = \frac{c}{f}.$$

(a) For the $f = 1530$ kHz AM radio signal, then,

Equation:

$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8 \text{ m/s}}{1530 \times 10^3 \text{ cycles/s}} \\ &= 196 \text{ m.}\end{aligned}$$

(b) For the $f = 105.1$ MHz FM radio signal,

Equation:

$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8 \text{ m/s}}{105.1 \times 10^6 \text{ cycles/s}} \\ &= 2.85 \text{ m.}\end{aligned}$$

(c) And for the $f = 1.90$ GHz cell phone,

Equation:

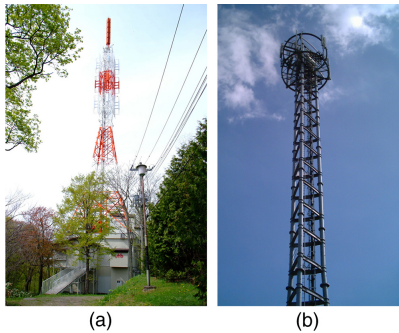
$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8 \text{ m/s}}{1.90 \times 10^9 \text{ cycles/s}} \\ &= 0.158 \text{ m.}\end{aligned}$$

Discussion

These wavelengths are consistent with the spectrum in [\[link\]](#). The wavelengths are also related to other properties of these electromagnetic waves, as we shall see.

The wavelengths found in the preceding example are representative of AM, FM, and cell phones, and account for some of the differences in how they are broadcast and how well they travel. The most efficient length for a linear antenna, such as discussed in [Production of Electromagnetic Waves](#), is $\lambda/2$, half the wavelength of the electromagnetic wave. Thus a very large antenna is needed to efficiently broadcast typical AM radio with its carrier wavelengths on the order of hundreds of meters.

One benefit to these long AM wavelengths is that they can go over and around rather large obstacles (like buildings and hills), just as ocean waves can go around large rocks. FM and TV are best received when there is a line of sight between the broadcast antenna and receiver, and they are often sent from very tall structures. FM, TV, and mobile phone antennas themselves are much smaller than those used for AM, but they are elevated to achieve an unobstructed line of sight. (See [\[link\]](#).)



(a) A large tower is used to broadcast TV signals.

The actual antennas are small structures on top of the tower—they are placed at great heights to have a clear line of sight over a large broadcast area. (credit: Ozizo, Wikimedia Commons)

(b) The NTT Dokomo mobile phone tower at Tokorozawa City, Japan.

(credit: tokoroten, Wikimedia Commons)

Radio Wave Interference

Astronomers and astrophysicists collect signals from outer space using electromagnetic waves. A common problem for astrophysicists is the “pollution” from electromagnetic radiation pervading our surroundings from communication systems in general. Even everyday gadgets like our car keys having the facility to lock car doors remotely and being able to turn TVs on and off using remotes involve radio-wave frequencies. In order to prevent interference between all these electromagnetic signals, strict regulations are drawn up for different organizations to utilize different radio frequency bands.

One reason why we are sometimes asked to switch off our mobile phones (operating in the range of 1.9 GHz) on airplanes and in hospitals is that important communications or medical equipment often uses similar radio frequencies and their operation can be affected by frequencies used in the communication devices.

For example, radio waves used in magnetic resonance imaging (MRI) have frequencies on the order of 100 MHz, although this varies significantly depending on the strength of the magnetic field used and the nuclear type being scanned. MRI is an important medical imaging and research tool, producing highly detailed two- and three-dimensional images. Radio waves are broadcast, absorbed, and reemitted in a resonance process that is sensitive to the density of nuclei (usually protons or hydrogen nuclei).

The wavelength of 100-MHz radio waves is 3 m, yet using the sensitivity of the resonant frequency to the magnetic field strength, details smaller than a millimeter can be imaged. This is a good example of an exception to a rule of thumb (in this case, the rubric that details much smaller than the probe's wavelength cannot be detected). The intensity of the radio waves used in MRI presents little or no hazard to human health.

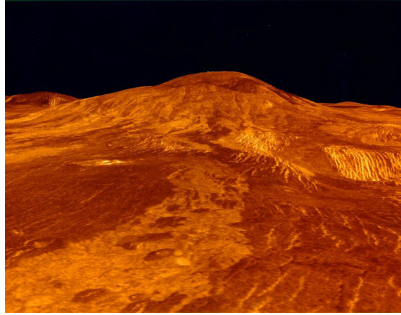
Microwaves

Microwaves are the highest-frequency electromagnetic waves that can be produced by currents in macroscopic circuits and devices. Microwave frequencies range from about 10^9 Hz to the highest practical LC resonance at nearly 10^{12} Hz. Since they have high frequencies, their wavelengths are short compared with those of other radio waves—hence the name “microwave.”

Microwaves can also be produced by atoms and molecules. They are, for example, a component of electromagnetic radiation generated by **thermal agitation**. The thermal motion of atoms and molecules in any object at a temperature above absolute zero causes them to emit and absorb radiation.

Since it is possible to carry more information per unit time on high frequencies, microwaves are quite suitable for communications. Most satellite-transmitted information is carried on microwaves, as are land-based long-distance transmissions. A clear line of sight between transmitter and receiver is needed because of the short wavelengths involved.

Radar is a common application of microwaves that was first developed in World War II. By detecting and timing microwave echoes, radar systems can determine the distance to objects as diverse as clouds and aircraft. A Doppler shift in the radar echo can be used to determine the speed of a car or the intensity of a rainstorm. Sophisticated radar systems are used to map the Earth and other planets, with a resolution limited by wavelength. (See [\[link\]](#).) The shorter the wavelength of any probe, the smaller the detail it is possible to observe.



An image of Sif Mons with lava flows on Venus, based on Magellan synthetic aperture radar data combined with radar altimetry to produce a three-dimensional map of the surface. The Venusian atmosphere is opaque to visible light, but not to the microwaves that were used to create this image.

(credit: NSSDC,
NASA/JPL)

Heating with Microwaves

How does the ubiquitous microwave oven produce microwaves electronically, and why does food absorb them preferentially? Microwaves at a frequency of 2.45 GHz are produced by accelerating electrons. The microwaves are then used to induce an alternating electric field in the oven.

Water and some other constituents of food have a slightly negative charge at one end and a slightly positive charge at one end (called polar molecules). The range of microwave frequencies is specially selected so that the polar molecules, in trying to keep orienting themselves with the electric field, absorb these energies and increase their temperatures—called dielectric heating.

The energy thereby absorbed results in thermal agitation heating food and not the plate, which does not contain water. Hot spots in the food are related to constructive and destructive interference patterns. Rotating antennas and food turntables help spread out the hot spots.

Another use of microwaves for heating is within the human body. Microwaves will penetrate more than shorter wavelengths into tissue and so can accomplish “deep heating” (called

microwave diathermy). This is used for treating muscular pains, spasms, tendonitis, and rheumatoid arthritis.

Note:

Making Connections: Take-Home Experiment—Microwave Ovens

1. Look at the door of a microwave oven. Describe the structure of the door. Why is there a metal grid on the door? How does the size of the holes in the grid compare with the wavelengths of microwaves used in microwave ovens? What is this wavelength?
2. Place a glass of water (about 250 ml) in the microwave and heat it for 30 seconds. Measure the temperature gain (the ΔT). Assuming that the power output of the oven is 1000 W, calculate the efficiency of the heat-transfer process.
3. Remove the rotating turntable or moving plate and place a cup of water in several places along a line parallel with the opening. Heat for 30 seconds and measure the ΔT for each position. Do you see cases of destructive interference?

Microwaves generated by atoms and molecules far away in time and space can be received and detected by electronic circuits. Deep space acts like a blackbody with a 2.7 K temperature, radiating most of its energy in the microwave frequency range. In 1964, Penzias and Wilson detected this radiation and eventually recognized that it was the radiation of the Big Bang's cooled remnants.

Infrared Radiation

The microwave and infrared regions of the electromagnetic spectrum overlap (see [\[link\]](#)).

Infrared radiation is generally produced by thermal motion and the vibration and rotation of atoms and molecules. Electronic transitions in atoms and molecules can also produce infrared radiation.

The range of infrared frequencies extends up to the lower limit of visible light, just below red. In fact, infrared means “below red.” Frequencies at its upper limit are too high to be produced by accelerating electrons in circuits, but small systems, such as atoms and molecules, can vibrate fast enough to produce these waves.

Water molecules rotate and vibrate particularly well at infrared frequencies, emitting and absorbing them so efficiently that the emissivity for skin is $e = 0.97$ in the infrared. Night-vision scopes can detect the infrared emitted by various warm objects, including humans, and convert it to visible light.

We can examine radiant heat transfer from a house by using a camera capable of detecting infrared radiation. Reconnaissance satellites can detect buildings, vehicles, and even individual humans by their infrared emissions, whose power radiation is proportional to the fourth power of the absolute temperature. More mundanely, we use infrared lamps, some of which are called

quartz heaters, to preferentially warm us because we absorb infrared better than our surroundings.

The Sun radiates like a nearly perfect blackbody (that is, it has $e = 1$), with a 6000 K surface temperature. About half of the solar energy arriving at the Earth is in the infrared region, with most of the rest in the visible part of the spectrum, and a relatively small amount in the ultraviolet. On average, 50 percent of the incident solar energy is absorbed by the Earth.

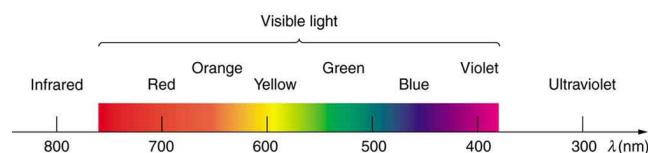
The relatively constant temperature of the Earth is a result of the energy balance between the incoming solar radiation and the energy radiated from the Earth. Most of the infrared radiation emitted from the Earth is absorbed by CO_2 and H_2O in the atmosphere and then radiated back to Earth or into outer space. This radiation back to Earth is known as the greenhouse effect, and it maintains the surface temperature of the Earth about 40°C higher than it would be if there is no absorption. Some scientists think that the increased concentration of CO_2 and other greenhouse gases in the atmosphere, resulting from increases in fossil fuel burning, has increased global average temperatures.

Visible Light

Visible light is the narrow segment of the electromagnetic spectrum to which the normal human eye responds. Visible light is produced by vibrations and rotations of atoms and molecules, as well as by electronic transitions within atoms and molecules. The receivers or detectors of light largely utilize electronic transitions. We say the atoms and molecules are excited when they absorb and relax when they emit through electronic transitions.

[\[link\]](#) shows this part of the spectrum, together with the colors associated with particular pure wavelengths. We usually refer to visible light as having wavelengths of between 400 nm and 750 nm. (The retina of the eye actually responds to the lowest ultraviolet frequencies, but these do not normally reach the retina because they are absorbed by the cornea and lens of the eye.)

Red light has the lowest frequencies and longest wavelengths, while violet has the highest frequencies and shortest wavelengths. Blackbody radiation from the Sun peaks in the visible part of the spectrum but is more intense in the red than in the violet, making the Sun yellowish in appearance.



A small part of the electromagnetic spectrum that includes its visible components. The divisions between infrared, visible, and ultraviolet are not perfectly

distinct, nor are those between the seven rainbow colors.

Living things—plants and animals—have evolved to utilize and respond to parts of the electromagnetic spectrum they are embedded in. Visible light is the most predominant and we enjoy the beauty of nature through visible light. Plants are more selective. Photosynthesis makes use of parts of the visible spectrum to make sugars.

Example:

Integrated Concept Problem: Correcting Vision with Lasers

During laser vision correction, a brief burst of 193-nm ultraviolet light is projected onto the cornea of a patient. It makes a spot 0.80 mm in diameter and evaporates a layer of cornea 0.30 μm thick. Calculate the energy absorbed, assuming the corneal tissue has the same properties as water; it is initially at 34°C. Assume the evaporated tissue leaves at a temperature of 100°C.

Strategy

The energy from the laser light goes toward raising the temperature of the tissue and also toward evaporating it. Thus we have two amounts of heat to add together. Also, we need to find the mass of corneal tissue involved.

Solution

To figure out the heat required to raise the temperature of the tissue to 100°C, we can apply concepts of thermal energy. We know that

Equation:

$$Q = mc\Delta T,$$

where Q is the heat required to raise the temperature, ΔT is the desired change in temperature, m is the mass of tissue to be heated, and c is the specific heat of water equal to 4186 J/kg/K. Without knowing the mass m at this point, we have

Equation:

$$Q = m(4186 \text{ J/kg/K})(100^\circ\text{C} - 34^\circ\text{C}) = m(276,276 \text{ J/kg}) = m(276 \text{ kJ/kg}).$$

The latent heat of vaporization of water is 2256 kJ/kg, so that the energy needed to evaporate mass m is

Equation:

$$Q_v = mL_v = m(2256 \text{ kJ/kg}).$$

To find the mass m , we use the equation $\rho = m/V$, where ρ is the density of the tissue and V is its volume. For this case,

Equation:

$$\begin{aligned}
 m &= \rho V \\
 &= (1000 \text{ kg/m}^3)(\text{area} \times \text{thickness}(\text{m}^3)) \\
 &= (1000 \text{ kg/m}^3)(\pi(0.80 \times 10^{-3} \text{ m})^2/4)(0.30 \times 10^{-6} \text{ m}) \\
 &= 0.151 \times 10^{-9} \text{ kg}.
 \end{aligned}$$

Therefore, the total energy absorbed by the tissue in the eye is the sum of Q and Q_v :

Equation:

$$Q_{\text{tot}} = m(c\Delta T + L_v) = (0.151 \times 10^{-9} \text{ kg})(276 \text{ kJ/kg} + 2256 \text{ kJ/kg}) = 382 \times 10^{-9} \text{ kJ}.$$

Discussion

The lasers used for this eye surgery are excimer lasers, whose light is well absorbed by biological tissue. They evaporate rather than burn the tissue, and can be used for precision work. Most lasers used for this type of eye surgery have an average power rating of about one watt. For our example, if we assume that each laser burst from this pulsed laser lasts for 10 ns, and there are 400 bursts per second, then the average power is $Q_{\text{tot}} \times 400 = 150 \text{ mW}$.

Optics is the study of the behavior of visible light and other forms of electromagnetic waves. Optics falls into two distinct categories. When electromagnetic radiation, such as visible light, interacts with objects that are large compared with its wavelength, its motion can be represented by straight lines like rays. Ray optics is the study of such situations and includes lenses and mirrors.

When electromagnetic radiation interacts with objects about the same size as the wavelength or smaller, its wave nature becomes apparent. For example, observable detail is limited by the wavelength, and so visible light can never detect individual atoms, because they are so much smaller than its wavelength. Physical or wave optics is the study of such situations and includes all wave characteristics.

Note:

Take-Home Experiment: Colors That Match

When you light a match you see largely orange light; when you light a gas stove you see blue light. Why are the colors different? What other colors are present in these?

Ultraviolet Radiation

Ultraviolet means “above violet.” The electromagnetic frequencies of **ultraviolet radiation (UV)** extend upward from violet, the highest-frequency visible light. Ultraviolet is also produced by atomic and molecular motions and electronic transitions. The wavelengths of ultraviolet extend from 400 nm down to about 10 nm at its highest frequencies, which overlap

with the lowest X-ray frequencies. It was recognized as early as 1801 by Johann Ritter that the solar spectrum had an invisible component beyond the violet range.

Solar UV radiation is broadly subdivided into three regions: UV-A (320–400 nm), UV-B (290–320 nm), and UV-C (220–290 nm), ranked from long to shorter wavelengths (from smaller to larger energies). Most UV-B and all UV-C is absorbed by ozone (O₃) molecules in the upper atmosphere. Consequently, 99% of the solar UV radiation reaching the Earth's surface is UV-A.

Human Exposure to UV Radiation

It is largely exposure to UV-B that causes skin cancer. It is estimated that as many as 20% of adults will develop skin cancer over the course of their lifetime. Again, treatment is often successful if caught early. Despite very little UV-B reaching the Earth's surface, there are substantial increases in skin-cancer rates in countries such as Australia, indicating how important it is that UV-B and UV-C continue to be absorbed by the upper atmosphere.

All UV radiation can damage collagen fibers, resulting in an acceleration of the aging process of skin and the formation of wrinkles. Because there is so little UV-B and UV-C reaching the Earth's surface, sunburn is caused by large exposures, and skin cancer from repeated exposure. Some studies indicate a link between overexposure to the Sun when young and melanoma later in life.

The tanning response is a defense mechanism in which the body produces pigments to absorb future exposures in inert skin layers above living cells. Basically UV-B radiation excites DNA molecules, distorting the DNA helix, leading to mutations and the possible formation of cancerous cells.

Repeated exposure to UV-B may also lead to the formation of cataracts in the eyes—a cause of blindness among people living in the equatorial belt where medical treatment is limited. Cataracts, clouding in the eye's lens and a loss of vision, are age related; 60% of those between the ages of 65 and 74 will develop cataracts. However, treatment is easy and successful, as one replaces the lens of the eye with a plastic lens. Prevention is important. Eye protection from UV is more effective with plastic sunglasses than those made of glass.

A major acute effect of extreme UV exposure is the suppression of the immune system, both locally and throughout the body.

Low-intensity ultraviolet is used to sterilize haircutting implements, implying that the energy associated with ultraviolet is deposited in a manner different from lower-frequency electromagnetic waves. (Actually this is true for all electromagnetic waves with frequencies greater than visible light.)

Flash photography is generally not allowed of precious artworks and colored prints because the UV radiation from the flash can cause photo-degradation in the artworks. Often artworks will have an extra-thick layer of glass in front of them, which is especially designed to absorb UV radiation.

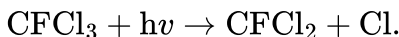
UV Light and the Ozone Layer

If all of the Sun's ultraviolet radiation reached the Earth's surface, there would be extremely grave effects on the biosphere from the severe cell damage it causes. However, the layer of ozone (O₃) in our upper atmosphere (10 to 50 km above the Earth) protects life by absorbing most of the dangerous UV radiation.

Unfortunately, today we are observing a depletion in ozone concentrations in the upper atmosphere. This depletion has led to the formation of an "ozone hole" in the upper atmosphere. The hole is more centered over the southern hemisphere, and changes with the seasons, being largest in the spring. This depletion is attributed to the breakdown of ozone molecules by refrigerant gases called chlorofluorocarbons (CFCs).

The UV radiation helps dissociate the CFC's, releasing highly reactive chlorine (Cl) atoms, which catalyze the destruction of the ozone layer. For example, the reaction of CFC₃ with a photon of light (hν) can be written as:

Equation:



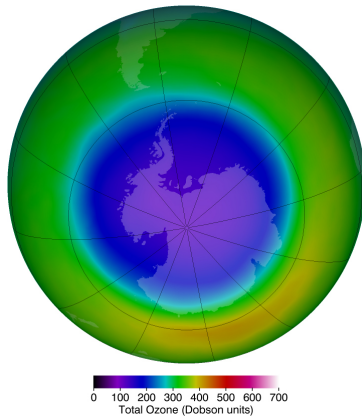
The Cl atom then catalyzes the breakdown of ozone as follows:

Equation:



A single chlorine atom could destroy ozone molecules for up to two years before being transported down to the surface. The CFCs are relatively stable and will contribute to ozone depletion for years to come. CFCs are found in refrigerants, air conditioning systems, foams, and aerosols.

International concern over this problem led to the establishment of the "Montreal Protocol" agreement (1987) to phase out CFC production in most countries. However, developing-country participation is needed if worldwide production and elimination of CFCs is to be achieved. Probably the largest contributor to CFC emissions today is India. But the protocol seems to be working, as there are signs of an ozone recovery. (See [\[link\]](#).)



This map of ozone concentration over Antarctica in October 2011 shows severe depletion suspected to be caused by CFCs.

Less dramatic but more general depletion has been observed over northern latitudes, suggesting the effect is global. With less ozone, more ultraviolet radiation from the Sun reaches the surface, causing more damage. (credit: NASA Ozone Watch)

Benefits of UV Light

Besides the adverse effects of ultraviolet radiation, there are also benefits of exposure in nature and uses in technology. Vitamin D production in the skin (epidermis) results from exposure to UVB radiation, generally from sunlight. A number of studies indicate lack of vitamin D can result in the development of a range of cancers (prostate, breast, colon), so a certain amount of UV exposure is helpful. Lack of vitamin D is also linked to osteoporosis. Exposures (with no sunscreen) of 10 minutes a day to arms, face, and legs might be sufficient to provide the accepted dietary level. However, in the winter time north of about 37° latitude, most UVB gets blocked by the atmosphere.

UV radiation is used in the treatment of infantile jaundice and in some skin conditions. It is also used in sterilizing workspaces and tools, and killing germs in a wide range of applications. It is

also used as an analytical tool to identify substances.

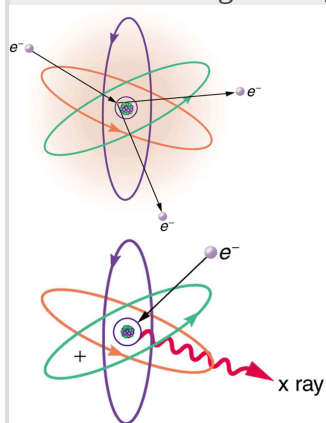
When exposed to ultraviolet, some substances, such as minerals, glow in characteristic visible wavelengths, a process called fluorescence. So-called black lights emit ultraviolet to cause posters and clothing to fluoresce in the visible. Ultraviolet is also used in special microscopes to detect details smaller than those observable with longer-wavelength visible-light microscopes.

Note:

Things Great and Small: A Submicroscopic View of X-Ray Production

X-rays can be created in a high-voltage discharge. They are emitted in the material struck by electrons in the discharge current. There are two mechanisms by which the electrons create X-rays.

The first method is illustrated in [\[link\]](#). An electron is accelerated in an evacuated tube by a high positive voltage. The electron strikes a metal plate (e.g., copper) and produces X-rays. Since this is a high-voltage discharge, the electron gains sufficient energy to ionize the atom.

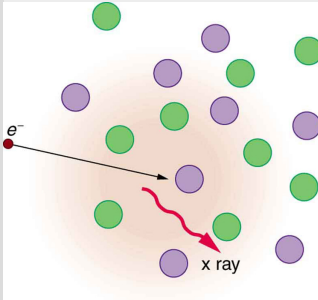


Artist's conception
of an electron
ionizing an atom
followed by the
recapture of an
electron and
emission of an X-
ray. An energetic
electron strikes an
atom and knocks an
electron out of one
of the orbits closest
to the nucleus.
Later, the atom
captures another
electron, and the
energy released by

its fall into a low orbit generates a high-energy EM wave called an X-ray.

In the case shown, an inner-shell electron (one in an orbit relatively close to and tightly bound to the nucleus) is ejected. A short time later, another electron is captured and falls into the orbit in a single great plunge. The energy released by this fall is given to an EM wave known as an X-ray. Since the orbits of the atom are unique to the type of atom, the energy of the X-ray is characteristic of the atom, hence the name characteristic X-ray.

The second method by which an energetic electron creates an X-ray when it strikes a material is illustrated in [\[link\]](#). The electron interacts with charges in the material as it penetrates. These collisions transfer kinetic energy from the electron to the electrons and atoms in the material.



Artist's conception of an electron being slowed by collisions in a material and emitting X-ray radiation. This energetic electron makes numerous collisions with electrons and atoms in a material it penetrates. An accelerated charge radiates EM waves, a second method by which X-rays are created.

A loss of kinetic energy implies an acceleration, in this case decreasing the electron's velocity. Whenever a charge is accelerated, it radiates EM waves. Given the high energy of the electron,

these EM waves can have high energy. We call them X-rays. Since the process is random, a broad spectrum of X-ray energy is emitted that is more characteristic of the electron energy than the type of material the electron encounters. Such EM radiation is called “bremsstrahlung” (German for “braking radiation”).

X-Rays

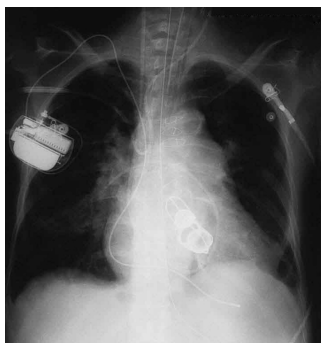
In the 1850s, scientists (such as Faraday) began experimenting with high-voltage electrical discharges in tubes filled with rarefied gases. It was later found that these discharges created an invisible, penetrating form of very high frequency electromagnetic radiation. This radiation was called an **X-ray**, because its identity and nature were unknown.

As described in [Things Great and Small](#), there are two methods by which X-rays are created—both are submicroscopic processes and can be caused by high-voltage discharges. While the low-frequency end of the X-ray range overlaps with the ultraviolet, X-rays extend to much higher frequencies (and energies).

X-rays have adverse effects on living cells similar to those of ultraviolet radiation, and they have the additional liability of being more penetrating, affecting more than the surface layers of cells. Cancer and genetic defects can be induced by exposure to X-rays. Because of their effect on rapidly dividing cells, X-rays can also be used to treat and even cure cancer.

The widest use of X-rays is for imaging objects that are opaque to visible light, such as the human body or aircraft parts. In humans, the risk of cell damage is weighed carefully against the benefit of the diagnostic information obtained. However, questions have risen in recent years as to accidental overexposure of some people during CT scans—a mistake at least in part due to poor monitoring of radiation dose.

The ability of X-rays to penetrate matter depends on density, and so an X-ray image can reveal very detailed density information. [\[link\]](#) shows an example of the simplest type of X-ray image, an X-ray shadow on film. The amount of information in a simple X-ray image is impressive, but more sophisticated techniques, such as CT scans, can reveal three-dimensional information with details smaller than a millimeter.



This shadow X-ray
image shows many
interesting features,
such as artificial
heart valves, a
pacemaker, and the
wires used to close
the sternum.
(credit: P. P. Urone)

The use of X-ray technology in medicine is called radiology—an established and relatively cheap tool in comparison to more sophisticated technologies. Consequently, X-rays are widely available and used extensively in medical diagnostics. During World War I, mobile X-ray units, advocated by Madame Marie Curie, were used to diagnose soldiers.

Because they can have wavelengths less than 0.01 nm, X-rays can be scattered (a process called X-ray diffraction) to detect the shape of molecules and the structure of crystals. X-ray diffraction was crucial to Crick, Watson, and Wilkins in the determination of the shape of the double-helix DNA molecule.

X-rays are also used as a precise tool for trace-metal analysis in X-ray induced fluorescence, in which the energy of the X-ray emissions are related to the specific types of elements and amounts of materials present.

Gamma Rays

Soon after nuclear radioactivity was first detected in 1896, it was found that at least three distinct types of radiation were being emitted. The most penetrating nuclear radiation was called a **gamma ray (γ ray)** (again a name given because its identity and character were unknown), and it was later found to be an extremely high frequency electromagnetic wave.

In fact, γ rays are any electromagnetic radiation emitted by a nucleus. This can be from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the γ -ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation.

Gamma rays have characteristics identical to X-rays of the same frequency—they differ only in source. At higher frequencies, γ rays are more penetrating and more damaging to living tissue. They have many of the same uses as X-rays, including cancer therapy. Gamma radiation from radioactive materials is used in nuclear medicine.

[\[link\]](#) shows a medical image based on γ rays. Food spoilage can be greatly inhibited by exposing it to large doses of γ radiation, thereby obliterating responsible microorganisms. Damage to food cells through irradiation occurs as well, and the long-term hazards of

consuming radiation-preserved food are unknown and controversial for some groups. Both X-ray and γ -ray technologies are also used in scanning luggage at airports.



This is an image of the γ rays emitted by nuclei in a compound that is concentrated in the bones and eliminated through the kidneys. Bone cancer is evidenced by nonuniform concentration in similar

structures.
For example,
some ribs are
darker than
others.
(credit: P. P.
Urone)

Detecting Electromagnetic Waves from Space

A final note on star gazing. The entire electromagnetic spectrum is used by researchers for investigating stars, space, and time. As noted earlier, Penzias and Wilson detected microwaves to identify the background radiation originating from the Big Bang. Radio telescopes such as the Arecibo Radio Telescope in Puerto Rico and Parkes Observatory in Australia were designed to detect radio waves.

Infrared telescopes need to have their detectors cooled by liquid nitrogen to be able to gather useful signals. Since infrared radiation is predominantly from thermal agitation, if the detectors were not cooled, the vibrations of the molecules in the antenna would be stronger than the signal being collected.

The most famous of these infrared sensitive telescopes is the James Clerk Maxwell Telescope in Hawaii. The earliest telescopes, developed in the seventeenth century, were optical telescopes, collecting visible light. Telescopes in the ultraviolet, X-ray, and γ -ray regions are placed outside the atmosphere on satellites orbiting the Earth.

The Hubble Space Telescope (launched in 1990) gathers ultraviolet radiation as well as visible light. In the X-ray region, there is the Chandra X-ray Observatory (launched in 1999), and in the γ -ray region, there is the new Fermi Gamma-ray Space Telescope (launched in 2008—taking the place of the Compton Gamma Ray Observatory, 1991–2000.).

Note:

PhET Explorations: Color Vision

Make a whole rainbow by mixing red, green, and blue light. Change the wavelength of a monochromatic beam or filter white light. View the light as a solid beam, or see the individual photons.

[Color
Vision](#)
[n](#)

Section Summary

- The relationship among the speed of propagation, wavelength, and frequency for any wave is given by $v_W = f\lambda$, so that for electromagnetic waves,

Equation:

$$c = f\lambda,$$

where f is the frequency, λ is the wavelength, and c is the speed of light.

- The electromagnetic spectrum is separated into many categories and subcategories, based on the frequency and wavelength, source, and uses of the electromagnetic waves.
- Any electromagnetic wave produced by currents in wires is classified as a radio wave, the lowest frequency electromagnetic waves. Radio waves are divided into many types, depending on their applications, ranging up to microwaves at their highest frequencies.
- Infrared radiation lies below visible light in frequency and is produced by thermal motion and the vibration and rotation of atoms and molecules. Infrared's lower frequencies overlap with the highest-frequency microwaves.
- Visible light is largely produced by electronic transitions in atoms and molecules, and is defined as being detectable by the human eye. Its colors vary with frequency, from red at the lowest to violet at the highest.
- Ultraviolet radiation starts with frequencies just above violet in the visible range and is produced primarily by electronic transitions in atoms and molecules.
- X-rays are created in high-voltage discharges and by electron bombardment of metal targets. Their lowest frequencies overlap the ultraviolet range but extend to much higher values, overlapping at the high end with gamma rays.
- Gamma rays are nuclear in origin and are defined to include the highest-frequency electromagnetic radiation of any type.

Conceptual Questions

Exercise:

Problem:

If you live in a region that has a particular TV station, you can sometimes pick up some of its audio portion on your FM radio receiver. Explain how this is possible. Does it imply that TV audio is broadcast as FM?

Exercise:

Problem:

Explain why people who have the lens of their eye removed because of cataracts are able to see low-frequency ultraviolet.

Exercise:

Problem:

How do fluorescent soap residues make clothing look “brighter and whiter” in outdoor light? Would this be effective in candlelight?

Exercise:

Problem: Give an example of resonance in the reception of electromagnetic waves.

Exercise:

Problem:

Illustrate that the size of details of an object that can be detected with electromagnetic waves is related to their wavelength, by comparing details observable with two different types (for example, radar and visible light or infrared and X-rays).

Exercise:

Problem: Why don’t buildings block radio waves as completely as they do visible light?

Exercise:

Problem:

Make a list of some everyday objects and decide whether they are transparent or opaque to each of the types of electromagnetic waves.

Exercise:

Problem:

Your friend says that more patterns and colors can be seen on the wings of birds if viewed in ultraviolet light. Would you agree with your friend? Explain your answer.

Exercise:

Problem:

The rate at which information can be transmitted on an electromagnetic wave is proportional to the frequency of the wave. Is this consistent with the fact that laser telephone transmission at visible frequencies carries far more conversations per optical fiber than conventional electronic transmission in a wire? What is the implication for ELF radio communication with submarines?

Exercise:

Problem: Give an example of energy carried by an electromagnetic wave.

Exercise:

Problem:

In an MRI scan, a higher magnetic field requires higher frequency radio waves to resonate with the nuclear type whose density and location is being imaged. What effect does going to a larger magnetic field have on the most efficient antenna to broadcast those radio waves? Does it favor a smaller or larger antenna?

Exercise:**Problem:**

Laser vision correction often uses an excimer laser that produces 193-nm electromagnetic radiation. This wavelength is extremely strongly absorbed by the cornea and ablates it in a manner that reshapes the cornea to correct vision defects. Explain how the strong absorption helps concentrate the energy in a thin layer and thus give greater accuracy in shaping the cornea. Also explain how this strong absorption limits damage to the lens and retina of the eye.

Problems & Exercises**Exercise:****Problem:**

(a) Two microwave frequencies are authorized for use in microwave ovens: 900 and 2560 MHz. Calculate the wavelength of each. (b) Which frequency would produce smaller hot spots in foods due to interference effects?

Solution:

(a) 33.3 cm (900 MHz) 11.7 cm (2560 MHz)

(b) The microwave oven with the smaller wavelength would produce smaller hot spots in foods, corresponding to the one with the frequency 2560 MHz.

Exercise:**Problem:**

(a) Calculate the range of wavelengths for AM radio given its frequency range is 540 to 1600 kHz. (b) Do the same for the FM frequency range of 88.0 to 108 MHz.

Exercise:**Problem:**

A radio station utilizes frequencies between commercial AM and FM. What is the frequency of a 11.12-m-wavelength channel?

Solution:

26.96 MHz

Exercise:

Problem:

Find the frequency range of visible light, given that it encompasses wavelengths from 380 to 760 nm.

Exercise:

Problem:

Combing your hair leads to excess electrons on the comb. How fast would you have to move the comb up and down to produce red light?

Solution:

$$5.0 \times 10^{14} \text{ Hz}$$

Exercise:

Problem:

Electromagnetic radiation having a $15.0 - \mu\text{m}$ wavelength is classified as infrared radiation. What is its frequency?

Exercise:

Problem:

Approximately what is the smallest detail observable with a microscope that uses ultraviolet light of frequency $1.20 \times 10^{15} \text{ Hz}$?

Solution:

Equation:

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^8 \text{ m/s}}{1.20 \times 10^{15} \text{ Hz}} = 2.50 \times 10^{-7} \text{ m}$$

Exercise:

Problem:

A radar used to detect the presence of aircraft receives a pulse that has reflected off an object $6 \times 10^{-5} \text{ s}$ after it was transmitted. What is the distance from the radar station to the reflecting object?

Exercise:

Problem:

Some radar systems detect the size and shape of objects such as aircraft and geological terrain. Approximately what is the smallest observable detail utilizing 500-MHz radar?

Solution:

0.600 m

Exercise:**Problem:**

Determine the amount of time it takes for X-rays of frequency 3×10^{18} Hz to travel (a) 1 mm and (b) 1 cm.

Exercise:**Problem:**

If you wish to detect details of the size of atoms (about 1×10^{-10} m) with electromagnetic radiation, it must have a wavelength of about this size. (a) What is its frequency? (b) What type of electromagnetic radiation might this be?

Solution:

$$(a) f = \frac{c}{\lambda} = \frac{3.00 \times 10^8 \text{ m/s}}{1 \times 10^{-10} \text{ m}} = 3 \times 10^{18} \text{ Hz}$$

(b) X-rays

Exercise:**Problem:**

If the Sun suddenly turned off, we would not know it until its light stopped coming. How long would that be, given that the Sun is 1.50×10^{11} m away?

Exercise:**Problem:**

Distances in space are often quoted in units of light years, the distance light travels in one year. (a) How many meters is a light year? (b) How many meters is it to Andromeda, the nearest large galaxy, given that it is 2.00×10^6 light years away? (c) The most distant galaxy yet discovered is 12.0×10^9 light years away. How far is this in meters?

Exercise:

Problem:

A certain 50.0-Hz AC power line radiates an electromagnetic wave having a maximum electric field strength of 13.0 kV/m. (a) What is the wavelength of this very low frequency electromagnetic wave? (b) What is its maximum magnetic field strength?

Solution:

(a) $6.00 \times 10^6 \text{ m}$

(b) $4.33 \times 10^{-5} \text{ T}$

Exercise:**Problem:**

During normal beating, the heart creates a maximum 4.00-mV potential across 0.300 m of a person's chest, creating a 1.00-Hz electromagnetic wave. (a) What is the maximum electric field strength created? (b) What is the corresponding maximum magnetic field strength in the electromagnetic wave? (c) What is the wavelength of the electromagnetic wave?

Exercise:**Problem:**

(a) The ideal size (most efficient) for a broadcast antenna with one end on the ground is one-fourth the wavelength ($\lambda/4$) of the electromagnetic radiation being sent out. If a new radio station has such an antenna that is 50.0 m high, what frequency does it broadcast most efficiently? Is this in the AM or FM band? (b) Discuss the analogy of the fundamental resonant mode of an air column closed at one end to the resonance of currents on an antenna that is one-fourth their wavelength.

Solution:

(a) $1.50 \times 10^6 \text{ Hz}$, AM band

(b) The resonance of currents on an antenna that is $1/4$ their wavelength is analogous to the fundamental resonant mode of an air column closed at one end, since the tube also has a length equal to $1/4$ the wavelength of the fundamental oscillation.

Exercise:**Problem:**

(a) What is the wavelength of 100-MHz radio waves used in an MRI unit? (b) If the frequencies are swept over a ± 1.00 range centered on 100 MHz, what is the range of wavelengths broadcast?

Exercise:

Problem:

(a) What is the frequency of the 193-nm ultraviolet radiation used in laser eye surgery? (b) Assuming the accuracy with which this EM radiation can ablate the cornea is directly proportional to wavelength, how much more accurate can this UV be than the shortest visible wavelength of light?

Solution:

(a) $1.55 \times 10^{15} \text{ Hz}$

(b) The shortest wavelength of visible light is 380 nm, so that

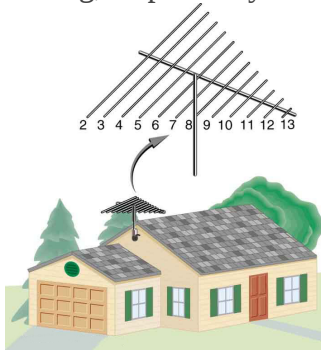
Equation:

$$\begin{aligned} & \frac{\lambda_{\text{visible}}}{\lambda_{\text{UV}}} \\ &= \frac{380 \text{ nm}}{193 \text{ nm}} \\ &= 1.97. \end{aligned}$$

In other words, the UV radiation is 97% more accurate than the shortest wavelength of visible light, or almost twice as accurate!

Exercise:**Problem:**

TV-reception antennas for VHF are constructed with cross wires supported at their centers, as shown in [\[link\]](#). The ideal length for the cross wires is one-half the wavelength to be received, with the more expensive antennas having one for each channel. Suppose you measure the lengths of the wires for particular channels and find them to be 1.94 and 0.753 m long, respectively. What are the frequencies for these channels?



A television reception antenna has cross wires of various lengths to most efficiently

receive different
wavelengths.

Exercise:

Problem:

Conversations with astronauts on lunar walks had an echo that was used to estimate the distance to the Moon. The sound spoken by the person on Earth was transformed into a radio signal sent to the Moon, and transformed back into sound on a speaker inside the astronaut's space suit. This sound was picked up by the microphone in the space suit (intended for the astronaut's voice) and sent back to Earth as a radio echo of sorts. If the round-trip time was 2.60 s, what was the approximate distance to the Moon, neglecting any delays in the electronics?

Solution:

$$3.90 \times 10^8 \text{ m}$$

Exercise:

Problem:

Lunar astronauts placed a reflector on the Moon's surface, off which a laser beam is periodically reflected. The distance to the Moon is calculated from the round-trip time. (a) To what accuracy in meters can the distance to the Moon be determined, if this time can be measured to 0.100 ns? (b) What percent accuracy is this, given the average distance to the Moon is $3.84 \times 10^8 \text{ m}$?

Exercise:

Problem:

Radar is used to determine distances to various objects by measuring the round-trip time for an echo from the object. (a) How far away is the planet Venus if the echo time is 1000 s? (b) What is the echo time for a car 75.0 m from a Highway Police radar unit? (c) How accurately (in nanoseconds) must you be able to measure the echo time to an airplane 12.0 km away to determine its distance within 10.0 m?

Solution:

(a) $1.50 \times 10^{11} \text{ m}$

(b) $0.500 \mu\text{s}$

(c) 66.7 ns

Exercise:

Problem: Integrated Concepts

- (a) Calculate the ratio of the highest to lowest frequencies of electromagnetic waves the eye can see, given the wavelength range of visible light is from 380 to 760 nm. (b) Compare this with the ratio of highest to lowest frequencies the ear can hear.

Exercise:**Problem: Integrated Concepts**

- (a) Calculate the rate in watts at which heat transfer through radiation occurs (almost entirely in the infrared) from 1.0 m^2 of the Earth's surface at night. Assume the emissivity is 0.90, the temperature of the Earth is 15°C , and that of outer space is 2.7 K. (b) Compare the intensity of this radiation with that coming to the Earth from the Sun during the day, which averages about 800 W/m^2 , only half of which is absorbed. (c) What is the maximum magnetic field strength in the outgoing radiation, assuming it is a continuous wave?

Solution:

- (a) $-3.5 \times 10^2 \text{ W/m}^2$
(b) 88%
(c) $1.7 \mu\text{T}$

Glossary

electromagnetic spectrum

the full range of wavelengths or frequencies of electromagnetic radiation

radio waves

electromagnetic waves with wavelengths in the range from 1 mm to 100 km; they are produced by currents in wires and circuits and by astronomical phenomena

microwaves

electromagnetic waves with wavelengths in the range from 1 mm to 1 m; they can be produced by currents in macroscopic circuits and devices

thermal agitation

the thermal motion of atoms and molecules in any object at a temperature above absolute zero, which causes them to emit and absorb radiation

radar

a common application of microwaves. Radar can determine the distance to objects as diverse as clouds and aircraft, as well as determine the speed of a car or the intensity of a

rainstorm

infrared radiation (IR)

a region of the electromagnetic spectrum with a frequency range that extends from just below the red region of the visible light spectrum up to the microwave region, or from $0.74\ \mu\text{m}$ to $300\ \mu\text{m}$

ultraviolet radiation (UV)

electromagnetic radiation in the range extending upward in frequency from violet light and overlapping with the lowest X-ray frequencies, with wavelengths from 400 nm down to about 10 nm

visible light

the narrow segment of the electromagnetic spectrum to which the normal human eye responds

amplitude modulation (AM)

a method for placing information on electromagnetic waves by modulating the amplitude of a carrier wave with an audio signal, resulting in a wave with constant frequency but varying amplitude

extremely low frequency (ELF)

electromagnetic radiation with wavelengths usually in the range of 0 to 300 Hz, but also about 1kHz

carrier wave

an electromagnetic wave that carries a signal by modulation of its amplitude or frequency

frequency modulation (FM)

a method of placing information on electromagnetic waves by modulating the frequency of a carrier wave with an audio signal, producing a wave of constant amplitude but varying frequency

TV

video and audio signals broadcast on electromagnetic waves

very high frequency (VHF)

TV channels utilizing frequencies in the two ranges of 54 to 88 MHz and 174 to 222 MHz

ultra-high frequency (UHF)

TV channels in an even higher frequency range than VHF, of 470 to 1000 MHz

X-ray

invisible, penetrating form of very high frequency electromagnetic radiation, overlapping both the ultraviolet range and the γ -ray range

gamma ray

(γ ray); extremely high frequency electromagnetic radiation emitted by the nucleus of an atom, either from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the γ -ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation

Energy in Electromagnetic Waves

- Explain how the energy and amplitude of an electromagnetic wave are related.
- Given its power output and the heating area, calculate the intensity of a microwave oven's electromagnetic field, as well as its peak electric and magnetic field strengths

Anyone who has used a microwave oven knows there is energy in **electromagnetic waves**. Sometimes this energy is obvious, such as in the warmth of the summer sun. Other times it is subtle, such as the unfelt energy of gamma rays, which can destroy living cells.

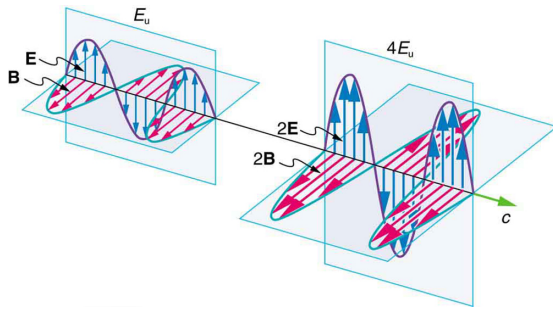
Electromagnetic waves can bring energy into a system by virtue of their **electric and magnetic fields**. These fields can exert forces and move charges in the system and, thus, do work on them. If the frequency of the electromagnetic wave is the same as the natural frequencies of the system (such as microwaves at the resonant frequency of water molecules), the transfer of energy is much more efficient.

Note:

Connections: Waves and Particles

The behavior of electromagnetic radiation clearly exhibits wave characteristics. But we shall find in later modules that at high frequencies, electromagnetic radiation also exhibits particle characteristics. These particle characteristics will be used to explain more of the properties of the electromagnetic spectrum and to introduce the formal study of modern physics.

Another startling discovery of modern physics is that particles, such as electrons and protons, exhibit wave characteristics. This simultaneous sharing of wave and particle properties for all submicroscopic entities is one of the great symmetries in nature.



Energy carried by a wave is proportional to its amplitude squared. With electromagnetic waves, larger E -fields and B -fields exert larger forces and can do more work.

But there is energy in an electromagnetic wave, whether it is absorbed or not. Once created, the fields carry energy away from a source. If absorbed, the field strengths are diminished and anything left travels on. Clearly, the larger the strength of the electric and magnetic fields, the more work they can do and the greater the energy the electromagnetic wave carries.

A wave's energy is proportional to its **amplitude** squared (E^2 or B^2). This is true for waves on guitar strings, for water waves, and for sound waves, where amplitude is proportional to pressure. In electromagnetic waves, the amplitude is the **maximum field strength** of the electric and magnetic fields. (See [\[link\]](#).)

Thus the energy carried and the **intensity** I of an electromagnetic wave is proportional to E^2 and B^2 . In fact, for a continuous sinusoidal electromagnetic wave, the average intensity I_{ave} is given by

Equation:

$$I_{\text{ave}} = \frac{c\epsilon_0 E_0^2}{2},$$

where c is the speed of light, ϵ_0 is the permittivity of free space, and E_0 is the maximum electric field strength; intensity, as always, is power per unit area (here in W/m^2).

The average intensity of an electromagnetic wave I_{ave} can also be expressed in terms of the magnetic field strength by using the relationship $B = E/c$, and the fact that $\epsilon_0 = 1/\mu_0 c^2$, where μ_0 is the permeability of free space. Algebraic manipulation produces the relationship

Equation:

$$I_{\text{ave}} = \frac{cB_0^2}{2\mu_0},$$

where B_0 is the maximum magnetic field strength.

One more expression for I_{ave} in terms of both electric and magnetic field strengths is useful. Substituting the fact that $c \cdot B_0 = E_0$, the previous expression becomes

Equation:

$$I_{\text{ave}} = \frac{E_0 B_0}{2\mu_0}.$$

Whichever of the three preceding equations is most convenient can be used, since they are really just different versions of the same principle: Energy in a wave is related to amplitude squared. Furthermore, since these equations are based on the assumption that the electromagnetic waves are sinusoidal, peak intensity is twice the average; that is, $I_0 = 2I_{\text{ave}}$.

Example:

Calculate Microwave Intensities and Fields

On its highest power setting, a certain microwave oven projects 1.00 kW of microwaves onto a 30.0 by 40.0 cm area. (a) What is the intensity in

W/m²? (b) Calculate the peak electric field strength E_0 in these waves.
(c) What is the peak magnetic field strength B_0 ?

Strategy

In part (a), we can find intensity from its definition as power per unit area. Once the intensity is known, we can use the equations below to find the field strengths asked for in parts (b) and (c).

Solution for (a)

Entering the given power into the definition of intensity, and noting the area is 0.300 by 0.400 m, yields

Equation:

$$I = \frac{P}{A} = \frac{1.00 \text{ kW}}{0.300 \text{ m} \times 0.400 \text{ m}}.$$

Here $I = I_{\text{ave}}$, so that

Equation:

$$I_{\text{ave}} = \frac{1000 \text{ W}}{0.120 \text{ m}^2} = 8.33 \times 10^3 \text{ W/m}^2.$$

Note that the peak intensity is twice the average:

Equation:

$$I_0 = 2I_{\text{ave}} = 1.67 \times 10^4 \text{ W/m}^2.$$

Solution for (b)

To find E_0 , we can rearrange the first equation given above for I_{ave} to give

Equation:

$$E_0 = \left(\frac{2I_{\text{ave}}}{c\epsilon_0} \right)^{1/2}.$$

Entering known values gives

Equation:

$$\begin{aligned}
 E_0 &= \sqrt{\frac{2(8.33 \times 10^3 \text{ W/m}^2)}{(3.00 \times 10^8 \text{ m/s})(8.85 \times 10^{-12} \text{ C}^2/\text{N}\cdot\text{m}^2)}} \\
 &= 2.51 \times 10^3 \text{ V/m}.
 \end{aligned}$$

Solution for (c)

Perhaps the easiest way to find magnetic field strength, now that the electric field strength is known, is to use the relationship given by

Equation:

$$B_0 = \frac{E_0}{c}.$$

Entering known values gives

Equation:

$$\begin{aligned}
 B_0 &= \frac{2.51 \times 10^3 \text{ V/m}}{3.0 \times 10^8 \text{ m/s}} \\
 &= 8.35 \times 10^{-6} \text{ T}.
 \end{aligned}$$

Discussion

As before, a relatively strong electric field is accompanied by a relatively weak magnetic field in an electromagnetic wave, since $B = E/c$, and c is a large number.

Section Summary

- The energy carried by any wave is proportional to its amplitude squared. For electromagnetic waves, this means intensity can be expressed as

Equation:

$$I_{\text{ave}} = \frac{c\varepsilon_0 E_0^2}{2},$$

where I_{ave} is the average intensity in W/m^2 , and E_0 is the maximum electric field strength of a continuous sinusoidal wave.

- This can also be expressed in terms of the maximum magnetic field strength B_0 as

Equation:

$$I_{\text{ave}} = \frac{cB_0^2}{2\mu_0}$$

and in terms of both electric and magnetic fields as

Equation:

$$I_{\text{ave}} = \frac{E_0 B_0}{2\mu_0}.$$

- The three expressions for I_{ave} are all equivalent.

Problems & Exercises

Exercise:

Problem:

What is the intensity of an electromagnetic wave with a peak electric field strength of 125 V/m?

Solution:

Equation:

$$\begin{aligned} I &= \frac{c\varepsilon_0 E_0^2}{2} \\ &= \frac{(3.00 \times 10^8 \text{ m/s})(8.85 \times 10^{-12} \text{ C}^2/\text{N}\cdot\text{m}^2)(125 \text{ V/m})^2}{2} \\ &= 20.7 \text{ W/m}^2 \end{aligned}$$

Exercise:

Problem:

Find the intensity of an electromagnetic wave having a peak magnetic field strength of $4.00 \times 10^{-9} \text{ T}$.

Exercise:**Problem:**

Assume the helium-neon lasers commonly used in student physics laboratories have power outputs of 0.250 mW. (a) If such a laser beam is projected onto a circular spot 1.00 mm in diameter, what is its intensity? (b) Find the peak magnetic field strength. (c) Find the peak electric field strength.

Solution:

$$(a) \ I = \frac{P}{A} = \frac{P}{\pi r^2} = \frac{0.250 \times 10^{-3} \text{ W}}{\pi (0.500 \times 10^{-3} \text{ m})^2} = 318 \text{ W/m}^2$$

$$\begin{aligned} I_{\text{ave}} &= \frac{cB_0^2}{2\mu_0} \Rightarrow B_0 = \left(\frac{2\mu_0 I}{c} \right)^{1/2} \\ (b) \quad &= \left(\frac{2(4\pi \times 10^{-7} \text{ T}\cdot\text{m/A})(318.3 \text{ W/m}^2)}{3.00 \times 10^8 \text{ m/s}} \right)^{1/2} \\ &= 1.63 \times 10^{-6} \text{ T} \end{aligned}$$

$$\begin{aligned} (c) \quad E_0 &= cB_0 = (3.00 \times 10^8 \text{ m/s})(1.633 \times 10^{-6} \text{ T}) \\ &= 4.90 \times 10^2 \text{ V/m} \end{aligned}$$

Exercise:

Problem:

An AM radio transmitter broadcasts 50.0 kW of power uniformly in all directions. (a) Assuming all of the radio waves that strike the ground are completely absorbed, and that there is no absorption by the atmosphere or other objects, what is the intensity 30.0 km away? (Hint: Half the power will be spread over the area of a hemisphere.) (b) What is the maximum electric field strength at this distance?

Exercise:**Problem:**

Suppose the maximum safe intensity of microwaves for human exposure is taken to be 1.00 W/m^2 . (a) If a radar unit leaks 10.0 W of microwaves (other than those sent by its antenna) uniformly in all directions, how far away must you be to be exposed to an intensity considered to be safe? Assume that the power spreads uniformly over the area of a sphere with no complications from absorption or reflection. (b) What is the maximum electric field strength at the safe intensity? (Note that early radar units leaked more than modern ones do. This caused identifiable health problems, such as cataracts, for people who worked near them.)

Solution:

(a) 89.2 cm

(b) 27.4 V/m

Exercise:

Problem:

A 2.50-m-diameter university communications satellite dish receives TV signals that have a maximum electric field strength (for one channel) of $7.50 \mu\text{V/m}$. (See [\[link\]](#).) (a) What is the intensity of this wave? (b) What is the power received by the antenna? (c) If the orbiting satellite broadcasts uniformly over an area of $1.50 \times 10^{13} \text{ m}^2$ (a large fraction of North America), how much power does it radiate?



Satellite dishes receive TV signals sent from orbit. Although the signals are quite weak, the receiver can detect them by being tuned to resonate at their frequency.

Exercise:

Problem:

Lasers can be constructed that produce an extremely high intensity electromagnetic wave for a brief time—called pulsed lasers. They are used to ignite nuclear fusion, for example. Such a laser may produce an electromagnetic wave with a maximum electric field strength of $1.00 \times 10^{11} \text{ V/m}$ for a time of 1.00 ns. (a) What is the maximum magnetic field strength in the wave? (b) What is the intensity of the beam? (c) What energy does it deliver on a 1.00-mm^2 area?

Solution:

(a) 333 T

(b) $1.33 \times 10^{19} \text{ W/m}^2$

(c) 13.3 kJ

Exercise:**Problem:**

Show that for a continuous sinusoidal electromagnetic wave, the peak intensity is twice the average intensity ($I_0 = 2I_{\text{ave}}$), using either the fact that $E_0 = \sqrt{2}E_{\text{rms}}$, or $B_0 = \sqrt{2}B_{\text{rms}}$, where rms means average (actually root mean square, a type of average).

Exercise:**Problem:**

Suppose a source of electromagnetic waves radiates uniformly in all directions in empty space where there are no absorption or interference effects. (a) Show that the intensity is inversely proportional to r^2 , the distance from the source squared. (b) Show that the magnitudes of the electric and magnetic fields are inversely proportional to r .

Solution:

$$(a) I = \frac{P}{A} = \frac{P}{4\pi r^2} \propto \frac{1}{r^2}$$

$$(b) I \propto E_0^2, B_0^2 \Rightarrow E_0^2, B_0^2 \propto \frac{1}{r^2} \Rightarrow E_0, B_0 \propto \frac{1}{r}$$

Exercise:

Problem: Integrated Concepts

An LC circuit with a 5.00-pF capacitor oscillates in such a manner as to radiate at a wavelength of 3.30 m. (a) What is the resonant frequency? (b) What inductance is in series with the capacitor?

Exercise:

Problem: Integrated Concepts

What capacitance is needed in series with an $800\text{ } \mu\text{H}$ inductor to form a circuit that radiates a wavelength of 196 m?

Solution:

13.5 pF

Exercise:

Problem: Integrated Concepts

Police radar determines the speed of motor vehicles using the same Doppler-shift technique employed for ultrasound in medical diagnostics. Beats are produced by mixing the double Doppler-shifted echo with the original frequency. If 1.50×10^9 -Hz microwaves are used and a beat frequency of 150 Hz is produced, what is the speed of the vehicle? (Assume the same Doppler-shift formulas are valid with the speed of sound replaced by the speed of light.)

Exercise:

Problem: Integrated Concepts

Assume the mostly infrared radiation from a heat lamp acts like a continuous wave with wavelength $1.50\ \mu\text{m}$. (a) If the lamp's 200-W output is focused on a person's shoulder, over a circular area 25.0 cm in diameter, what is the intensity in W/m^2 ? (b) What is the peak electric field strength? (c) Find the peak magnetic field strength. (d) How long will it take to increase the temperature of the 4.00-kg shoulder by 2.00°C , assuming no other heat transfer and given that its specific heat is $3.47 \times 10^3\ \text{J}/\text{kg}\cdot^\circ\text{C}$?

Solution:

(a) $4.07\ \text{kW}/\text{m}^2$

(b) $1.75\ \text{kV}/\text{m}$

(c) $5.84\ \mu\text{T}$

(d) 2 min 19 s

Exercise:

Problem: Integrated Concepts

On its highest power setting, a microwave oven increases the temperature of 0.400 kg of spaghetti by 45.0°C in 120 s. (a) What was the rate of power absorption by the spaghetti, given that its specific heat is $3.76 \times 10^3\ \text{J}/\text{kg}\cdot^\circ\text{C}$? (b) Find the average intensity of the microwaves, given that they are absorbed over a circular area 20.0 cm in diameter. (c) What is the peak electric field strength of the microwave? (d) What is its peak magnetic field strength?

Exercise:

Problem: Integrated Concepts

Electromagnetic radiation from a 5.00-mW laser is concentrated on a 1.00-mm^2 area. (a) What is the intensity in W/m^2 ? (b) Suppose a 2.00-nC static charge is in the beam. What is the maximum electric

force it experiences? (c) If the static charge moves at 400 m/s, what maximum magnetic force can it feel?

Solution:

(a) $5.00 \times 10^3 \text{ W/m}^2$

(b) $3.88 \times 10^{-6} \text{ N}$

(c) $5.18 \times 10^{-12} \text{ N}$

Exercise:

Problem: Integrated Concepts

A 200-turn flat coil of wire 30.0 cm in diameter acts as an antenna for FM radio at a frequency of 100 MHz. The magnetic field of the incoming electromagnetic wave is perpendicular to the coil and has a maximum strength of $1.00 \times 10^{-12} \text{ T}$. (a) What power is incident on the coil? (b) What average emf is induced in the coil over one-fourth of a cycle? (c) If the radio receiver has an inductance of $2.50 \mu\text{H}$, what capacitance must it have to resonate at 100 MHz?

Exercise:

Problem: Integrated Concepts

If electric and magnetic field strengths vary sinusoidally in time, being zero at $t = 0$, then $E = E_0 \sin 2\pi ft$ and $B = B_0 \sin 2\pi ft$. Let $f = 1.00 \text{ GHz}$ here. (a) When are the field strengths first zero? (b) When do they reach their most negative value? (c) How much time is needed for them to complete one cycle?

Solution:

(a) $t = 0$

(b) $7.50 \times 10^{-10} \text{ s}$

(c) $1.00 \times 10^{-9} \text{ s}$

Exercise:

Problem: Unreasonable Results

A researcher measures the wavelength of a 1.20-GHz electromagnetic wave to be 0.500 m. (a) Calculate the speed at which this wave propagates. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: Unreasonable Results

The peak magnetic field strength in a residential microwave oven is $9.20 \times 10^{-5} \text{ T}$. (a) What is the intensity of the microwave? (b) What is unreasonable about this result? (c) What is wrong about the premise?

Solution:

(a) $1.01 \times 10^6 \text{ W/m}^2$

(b) Much too great for an oven.

(c) The assumed magnetic field is unreasonably large.

Exercise:

Problem: Unreasonable Results

An LC circuit containing a 2.00-H inductor oscillates at such a frequency that it radiates at a 1.00-m wavelength. (a) What is the capacitance of the circuit? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: Unreasonable Results

An LC circuit containing a 1.00-pF capacitor oscillates at such a frequency that it radiates at a 300-nm wavelength. (a) What is the inductance of the circuit? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $2.53 \times 10^{-20} \text{ H}$

(b) L is much too small.

(c) The wavelength is unreasonably small.

Exercise:

Problem: Create Your Own Problem

Consider electromagnetic fields produced by high voltage power lines. Construct a problem in which you calculate the intensity of this electromagnetic radiation in W/m^2 based on the measured magnetic field strength of the radiation in a home near the power lines. Assume these magnetic field strengths are known to average less than a μT . The intensity is small enough that it is difficult to imagine mechanisms for biological damage due to it. Discuss how much energy may be radiating from a section of power line several hundred meters long and compare this to the power likely to be carried by the lines. An idea of how much power this is can be obtained by calculating the approximate current responsible for μT fields at distances of tens of meters.

Exercise:

Problem: Create Your Own Problem

Consider the most recent generation of residential satellite dishes that are a little less than half a meter in diameter. Construct a problem in which you calculate the power received by the dish and the maximum electric field strength of the microwave signals for a single channel

received by the dish. Among the things to be considered are the power broadcast by the satellite and the area over which the power is spread, as well as the area of the receiving dish.

Glossary

maximum field strength

the maximum amplitude an electromagnetic wave can reach, representing the maximum amount of electric force and/or magnetic flux that the wave can exert

intensity

the power of an electric or magnetic field per unit area, for example, Watts per square meter

Introduction to Wave Optics

class="introduction"

The colors
reflected
by this
compact
disc vary
with angle
and are
not caused
by
pigments.

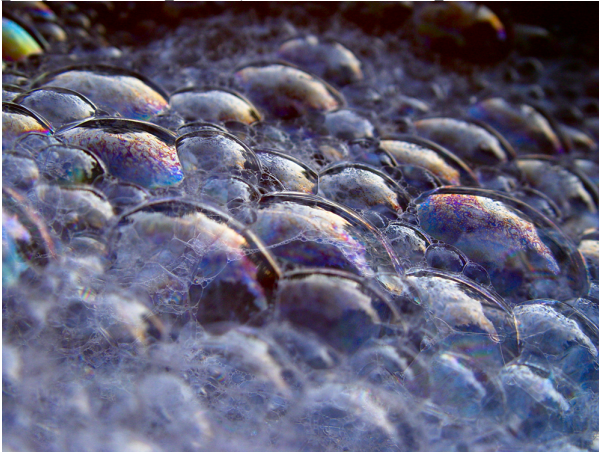
Colors
such as
these are
direct
evidence
of the
wave
character
of light.
(credit:
Infopro,
Wikimedi
a
Commons
)



Examine a compact disc under white light, noting the colors observed and locations of the colors. Determine if the spectra are formed by diffraction from circular lines centered at the middle of the disc and, if so, what is their spacing. If not, determine the type of spacing. Also with the CD, explore the spectra of a few light sources, such as a candle flame, incandescent bulb, halogen light, and fluorescent light. Knowing the spacing of the rows of pits in the compact disc, estimate the maximum spacing that will allow the given number of megabytes of information to be stored.

If you have ever looked at the reds, blues, and greens in a sunlit soap bubble and wondered how straw-colored soapy water could produce them, you have hit upon one of the many phenomena that can only be explained by the wave character of light (see [\[link\]](#)). The same is true for the colors seen in an oil slick or in the light reflected from a compact disc. These and other interesting phenomena, such as the dispersion of white light into a rainbow of colors when passed through a narrow slit, cannot be explained fully by geometric optics. In these cases, light interacts with small objects and exhibits its wave characteristics. The branch of optics that considers the

behavior of light when it exhibits wave characteristics (particularly when it interacts with small objects) is called wave optics (sometimes called physical optics). It is the topic of this chapter.



These soap bubbles exhibit brilliant colors when exposed to sunlight. How are the colors produced if they are not pigments in the soap? (credit: Scott Robinson, Flickr)

The Wave Aspect of Light: Interference

- Discuss the wave character of light.
- Identify the changes when light enters a medium.

We know that visible light is the type of electromagnetic wave to which our eyes respond. Like all other electromagnetic waves, it obeys the equation

Equation:

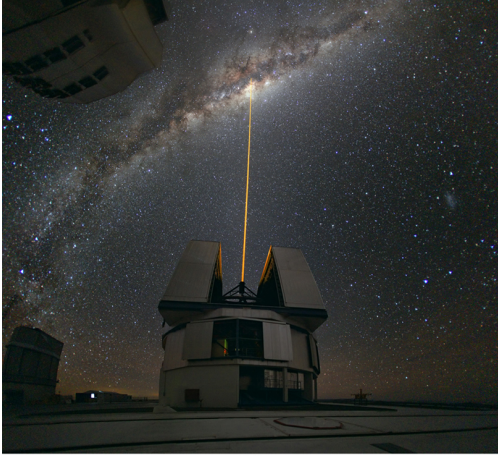
$$c = f\lambda,$$

where $c = 3 \times 10^8$ m/s is the speed of light in vacuum, f is the frequency of the electromagnetic waves, and λ is its wavelength. The range of visible wavelengths is approximately 380 to 760 nm. As is true for all waves, light travels in straight lines and acts like a ray when it interacts with objects several times as large as its wavelength. However, when it interacts with smaller objects, it displays its wave characteristics prominently. Interference is the hallmark of a wave, and in [\[link\]](#) both the ray and wave characteristics of light can be seen. The laser beam emitted by the observatory epitomizes a ray, traveling in a straight line. However, passing a pure-wavelength beam through vertical slits with a size close to the wavelength of the beam reveals the wave character of light, as the beam spreads out horizontally into a pattern of bright and dark regions caused by systematic constructive and destructive interference. Rather than spreading out, a ray would continue traveling straight ahead after passing through slits.

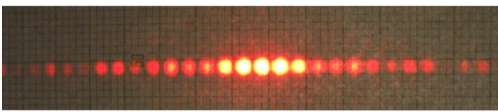
Note:

Making Connections: Waves

The most certain indication of a wave is interference. This wave characteristic is most prominent when the wave interacts with an object that is not large compared with the wavelength. Interference is observed for water waves, sound waves, light waves, and (as we will see in [Special Relativity](#)) for matter waves, such as electrons scattered from a crystal.



(a)



(b)

(a) The laser beam emitted by an observatory acts like a ray, traveling in a straight line. This laser beam is from the Paranal Observatory of the European Southern Observatory. (credit: Yuri Beletsky, European Southern Observatory)

(b) A laser beam passing through a grid of vertical slits produces an interference pattern—characteristic of a wave.

(credit: Shim'on and Slava Rybka, Wikimedia Commons)

Light has wave characteristics in various media as well as in a vacuum. When light goes from a vacuum to some medium, like water, its speed and

wavelength change, but its frequency f remains the same. (We can think of light as a forced oscillation that must have the frequency of the original source.) The speed of light in a medium is $v = c/n$, where n is its index of refraction. If we divide both sides of equation $c = f\lambda$ by n , we get $c/n = v = f\lambda/n$. This implies that $v = f\lambda_n$, where λ_n is the **wavelength in a medium** and that

Equation:

$$\lambda_n = \frac{\lambda}{n},$$

where λ is the wavelength in vacuum and n is the medium's index of refraction. Therefore, the wavelength of light is smaller in any medium than it is in vacuum. In water, for example, which has $n = 1.333$, the range of visible wavelengths is $(380 \text{ nm})/1.333$ to $(760 \text{ nm})/1.333$, or $\lambda_n = 285$ to 570 nm . Although wavelengths change while traveling from one medium to another, colors do not, since colors are associated with frequency.

Section Summary

- Wave optics is the branch of optics that must be used when light interacts with small objects or whenever the wave characteristics of light are considered.
- Wave characteristics are those associated with interference and diffraction.
- Visible light is the type of electromagnetic wave to which our eyes respond and has a wavelength in the range of 380 to 760 nm.
- Like all EM waves, the following relationship is valid in vacuum: $c = f\lambda$, where $c = 3 \times 10^8 \text{ m/s}$ is the speed of light, f is the frequency of the electromagnetic wave, and λ is its wavelength in vacuum.
- The wavelength λ_n of light in a medium with index of refraction n is $\lambda_n = \lambda/n$. Its frequency is the same as in vacuum.

Conceptual Questions

Exercise:

Problem:

What type of experimental evidence indicates that light is a wave?

Exercise:

Problem:

Give an example of a wave characteristic of light that is easily observed outside the laboratory.

Problems & Exercises

Exercise:

Problem:

Show that when light passes from air to water, its wavelength decreases to 0.750 times its original value.

Solution:

$$1/1.333 = 0.750$$

Exercise:

Problem:

Find the range of visible wavelengths of light in crown glass.

Exercise:

Problem:

What is the index of refraction of a material for which the wavelength of light is 0.671 times its value in a vacuum? Identify the likely substance.

Solution:

1.49, Polystyrene

Exercise:**Problem:**

Analysis of an interference effect in a clear solid shows that the wavelength of light in the solid is 329 nm. Knowing this light comes from a He-Ne laser and has a wavelength of 633 nm in air, is the substance zircon or diamond?

Exercise:**Problem:**

What is the ratio of thicknesses of crown glass and water that would contain the same number of wavelengths of light?

Solution:

0.877 glass to water

Glossary

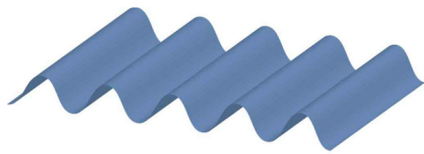
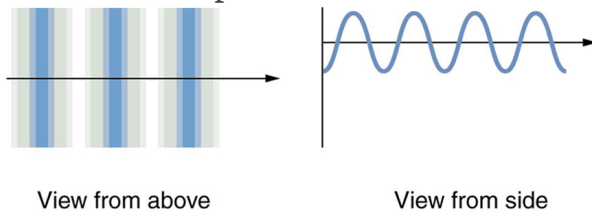
wavelength in a medium

$\lambda_n = \lambda/n$, where λ is the wavelength in vacuum, and n is the index of refraction of the medium

Huygens's Principle: Diffraction

- Discuss the propagation of transverse waves.
- Discuss Huygens's principle.
- Explain the bending of light.

[\[link\]](#) shows how a transverse wave looks as viewed from above and from the side. A light wave can be imagined to propagate like this, although we do not actually see it wiggling through space. From above, we view the wavefronts (or wave crests) as we would by looking down on the ocean waves. The side view would be a graph of the electric or magnetic field. The view from above is perhaps the most useful in developing concepts about wave optics.



Overall view

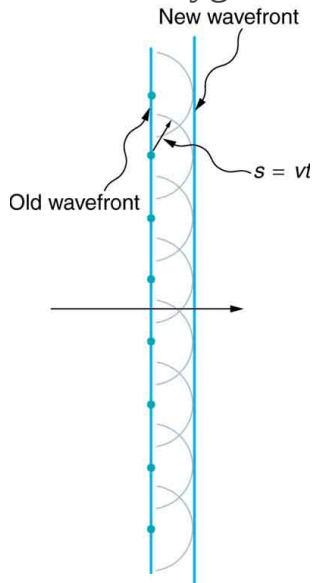
A transverse wave, such as an electromagnetic wave like light, as viewed from above and from the side. The direction of propagation is perpendicular to the wavefronts (or wave crests) and is represented by an arrow like a ray.

The Dutch scientist Christiaan Huygens (1629–1695) developed a useful technique for determining in detail how and where waves propagate.

Starting from some known position, **Huygens's principle** states that:

Every point on a wavefront is a source of wavelets that spread out in the forward direction at the same speed as the wave itself. The new wavefront is a line tangent to all of the wavelets.

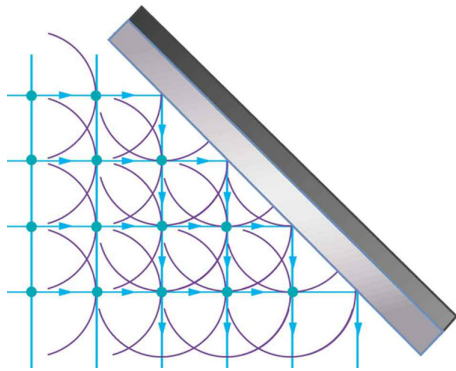
[\[link\]](#) shows how Huygens's principle is applied. A wavefront is the long edge that moves, for example, the crest or the trough. Each point on the wavefront emits a semicircular wave that moves at the propagation speed v . These are drawn at a time t later, so that they have moved a distance $s = vt$. The new wavefront is a line tangent to the wavelets and is where we would expect the wave to be a time t later. Huygens's principle works for all types of waves, including water waves, sound waves, and light waves. We will find it useful not only in describing how light waves propagate, but also in explaining the laws of reflection and refraction. In addition, we will see that Huygens's principle tells us how and where light rays interfere.



Huygens's
principle
applied to a
straight
wavefront.
Each point
on the
wavefront

emits a
semicircular
wavelet that
moves a
distance
 $s = vt$. The
new
wavefront is
a line tangent
to the
wavelets.

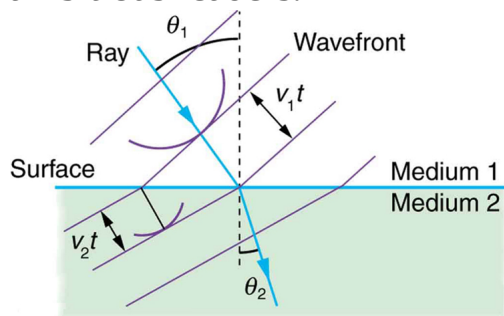
[\[link\]](#) shows how a mirror reflects an incoming wave at an angle equal to the incident angle, verifying the law of reflection. As the wavefront strikes the mirror, wavelets are first emitted from the left part of the mirror and then the right. The wavelets closer to the left have had time to travel farther, producing a wavefront traveling in the direction shown.



Huygens's principle
applied to a straight
wavefront striking a
mirror. The wavelets
shown were emitted as
each point on the
wavefront struck the
mirror. The tangent to
these wavelets shows

that the new wavefront has been reflected at an angle equal to the incident angle. The direction of propagation is perpendicular to the wavefront, as shown by the downward-pointing arrows.

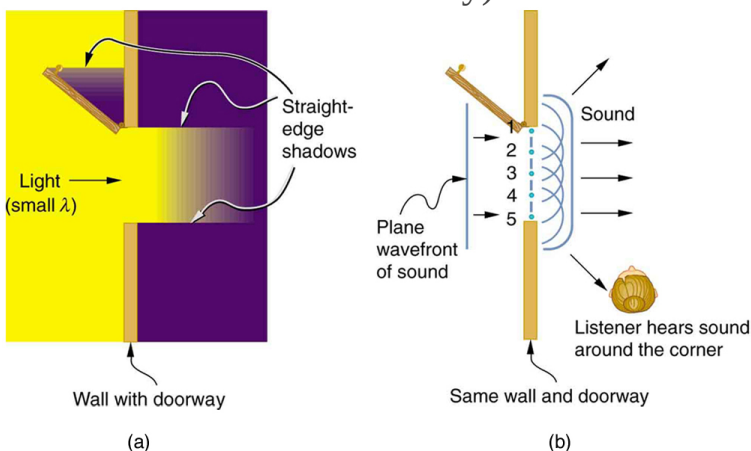
The law of refraction can be explained by applying Huygens's principle to a wavefront passing from one medium to another (see [\[link\]](#)). Each wavelet in the figure was emitted when the wavefront crossed the interface between the media. Since the speed of light is smaller in the second medium, the waves do not travel as far in a given time, and the new wavefront changes direction as shown. This explains why a ray changes direction to become closer to the perpendicular when light slows down. Snell's law can be derived from the geometry in [\[link\]](#), but this is left as an exercise for ambitious readers.



Huygens's principle applied to a straight wavefront traveling from one medium to another where its speed is less. The ray bends toward the perpendicular, since the

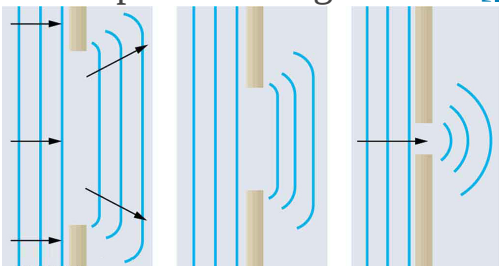
wavelets have a lower speed in the second medium.

What happens when a wave passes through an opening, such as light shining through an open door into a dark room? For light, we expect to see a sharp shadow of the doorway on the floor of the room, and we expect no light to bend around corners into other parts of the room. When sound passes through a door, we expect to hear it everywhere in the room and, thus, expect that sound spreads out when passing through such an opening (see [\[link\]](#)). What is the difference between the behavior of sound waves and light waves in this case? The answer is that light has very short wavelengths and acts like a ray. Sound has wavelengths on the order of the size of the door and bends around corners (for frequency of 1000 Hz, $\lambda = c/f = (330 \text{ m/s})/(1000 \text{ s}^{-1}) = 0.33 \text{ m}$, about three times smaller than the width of the doorway).



(a) Light passing through a doorway makes a sharp outline on the floor. Since light's wavelength is very small compared with the size of the door, it acts like a ray. (b) Sound waves bend into all parts of the room, a wave effect, because their wavelength is similar to the size of the door.

If we pass light through smaller openings, often called slits, we can use Huygens's principle to see that light bends as sound does (see [\[link\]](#)). The bending of a wave around the edges of an opening or an obstacle is called **diffraction**. Diffraction is a wave characteristic and occurs for all types of waves. If diffraction is observed for some phenomenon, it is evidence that the phenomenon is a wave. Thus the horizontal diffraction of the laser beam after it passes through slits in [\[link\]](#) is evidence that light is a wave.



Huygens's principle applied to a straight wavefront striking an opening. The edges of the wavefront bend after passing through the opening, a process called diffraction. The amount of bending is more extreme for a small opening, consistent with the fact that wave characteristics are most noticeable for interactions with objects about the same size as the wavelength.

Section Summary

- An accurate technique for determining how and where waves propagate is given by Huygens's principle: Every point on a wavefront is a source of wavelets that spread out in the forward direction at the same speed as the wave itself. The new wavefront is a line tangent to all of the wavelets.
- Diffraction is the bending of a wave around the edges of an opening or other obstacle.

Conceptual Questions

Exercise:

Problem:

How do wave effects depend on the size of the object with which the wave interacts? For example, why does sound bend around the corner of a building while light does not?

Exercise:

Problem:

Under what conditions can light be modeled like a ray? Like a wave?

Exercise:

Problem:

Go outside in the sunlight and observe your shadow. It has fuzzy edges even if you do not. Is this a diffraction effect? Explain.

Exercise:

Problem:

Why does the wavelength of light decrease when it passes from vacuum into a medium? State which attributes change and which stay the same and, thus, require the wavelength to decrease.

Exercise:

Problem: Does Huygens's principle apply to all types of waves?

Glossary

diffraction

the bending of a wave around the edges of an opening or an obstacle

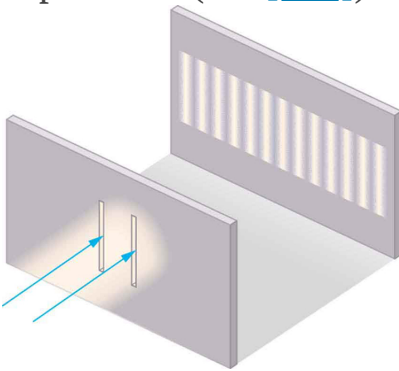
Huygens's principle

every point on a wavefront is a source of wavelets that spread out in the forward direction at the same speed as the wave itself. The new wavefront is a line tangent to all of the wavelets

Young's Double Slit Experiment

- Explain the phenomena of interference.
- Define constructive interference for a double slit and destructive interference for a double slit.

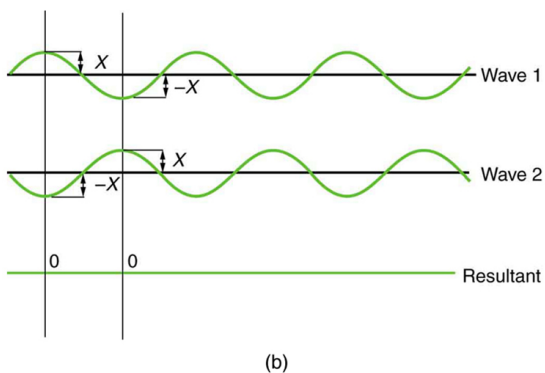
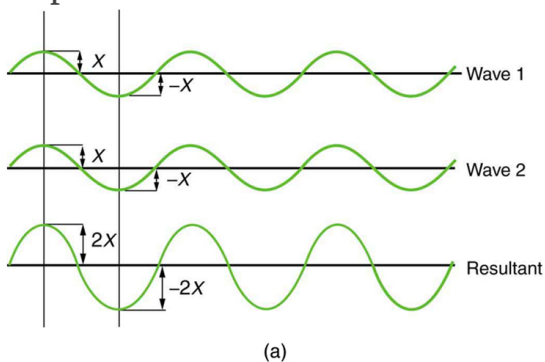
Although Christiaan Huygens thought that light was a wave, Isaac Newton did not. Newton felt that there were other explanations for color, and for the interference and diffraction effects that were observable at the time. Owing to Newton's tremendous stature, his view generally prevailed. The fact that Huygens's principle worked was not considered evidence that was direct enough to prove that light is a wave. The acceptance of the wave character of light came many years later when, in 1801, the English physicist and physician Thomas Young (1773–1829) did his now-classic double slit experiment (see [\[link\]](#)).



Young's double slit experiment. Here pure-wavelength light sent through a pair of vertical slits is diffracted into a pattern on the screen of numerous vertical lines spread out horizontally. Without diffraction and interference, the light would

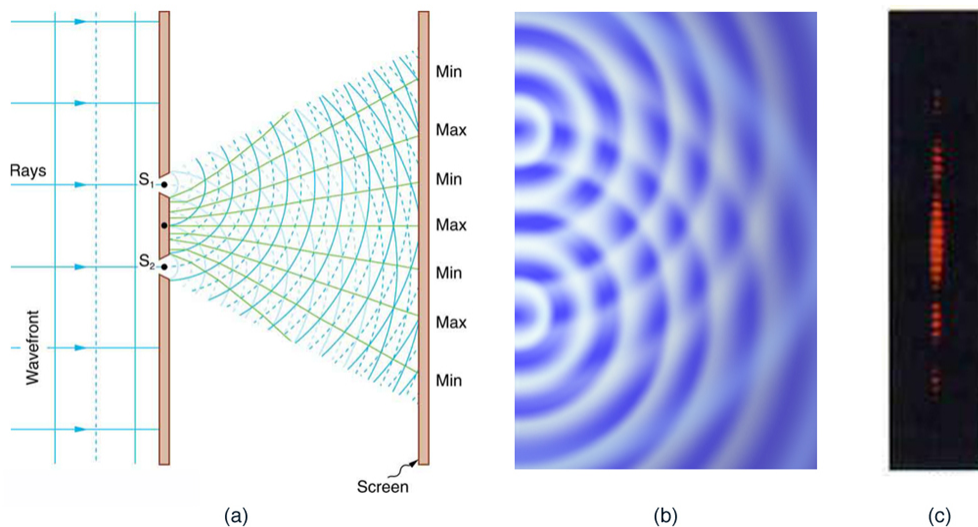
simply make two
lines on the screen.

Why do we not ordinarily observe wave behavior for light, such as observed in Young's double slit experiment? First, light must interact with something small, such as the closely spaced slits used by Young, to show pronounced wave effects. Furthermore, Young first passed light from a single source (the Sun) through a single slit to make the light somewhat coherent. By **coherent**, we mean waves are in phase or have a definite phase relationship. **Incoherent** means the waves have random phase relationships. Why did Young then pass the light through a double slit? The answer to this question is that two slits provide two coherent light sources that then interfere constructively or destructively. Young used sunlight, where each wavelength forms its own pattern, making the effect more difficult to see. We illustrate the double slit experiment with monochromatic (single λ) light to clarify the effect. [\[link\]](#) shows the pure constructive and destructive interference of two waves having the same wavelength and amplitude.



The amplitudes of waves add. (a) Pure constructive interference is obtained when identical waves are in phase. (b) Pure destructive interference occurs when identical waves are exactly out of phase, or shifted by half a wavelength.

When light passes through narrow slits, it is diffracted into semicircular waves, as shown in [\[link\]](#)(a). Pure constructive interference occurs where the waves are crest to crest or trough to trough. Pure destructive interference occurs where they are crest to trough. The light must fall on a screen and be scattered into our eyes for us to see the pattern. An analogous pattern for water waves is shown in [\[link\]](#)(b). Note that regions of constructive and destructive interference move out from the slits at well-defined angles to the original beam. These angles depend on wavelength and the distance between the slits, as we shall see below.



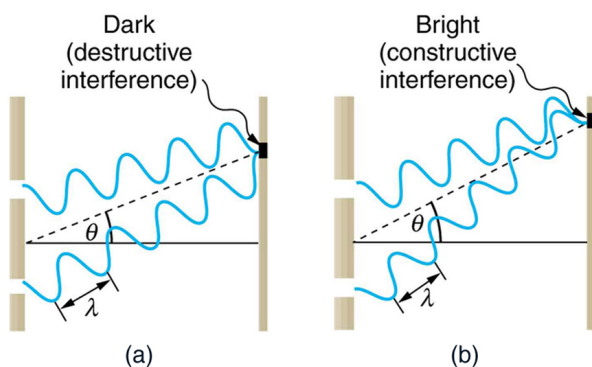
Double slits produce two coherent sources of waves that interfere. (a) Light spreads out (diffracts) from each slit, because the slits are narrow. These waves overlap and

interfere constructively (bright lines) and destructively (dark regions). We can only see this if the light falls onto a screen and is scattered into our eyes. (b) Double slit interference pattern for water waves are nearly identical to that for light. Wave action is greatest in regions of constructive interference and least in regions of destructive interference. (c) When light that has passed through double slits falls on a screen, we see a pattern such as this. (credit: PASCO)

To understand the double slit interference pattern, we consider how two waves travel from the slits to the screen, as illustrated in [\[link\]](#). Each slit is a different distance from a given point on the screen. Thus different numbers of wavelengths fit into each path. Waves start out from the slits in phase (crest to crest), but they may end up out of phase (crest to trough) at the screen if the paths differ in length by half a wavelength, interfering destructively as shown in [\[link\]](#)(a). If the paths differ by a whole wavelength, then the waves arrive in phase (crest to crest) at the screen, interfering constructively as shown in [\[link\]](#)(b). More generally, if the paths taken by the two waves differ by any half-integral number of wavelengths [$(1/2)\lambda$, $(3/2)\lambda$, $(5/2)\lambda$, etc.], then destructive interference occurs. Similarly, if the paths taken by the two waves differ by any integral number of wavelengths (λ , 2λ , 3λ , etc.), then constructive interference occurs.

Note:**Take-Home Experiment: Using Fingers as Slits**

Look at a light, such as a street lamp or incandescent bulb, through the narrow gap between two fingers held close together. What type of pattern do you see? How does it change when you allow the fingers to move a little farther apart? Is it more distinct for a monochromatic source, such as the yellow light from a sodium vapor lamp, than for an incandescent bulb?



Waves follow different paths from the slits to a common point on a screen. (a)

Destructive interference occurs here, because one path is a half wavelength longer than the other. The waves start in phase but arrive out of phase. (b)

Constructive interference occurs here because one path is a whole wavelength longer than the other. The waves start out and arrive in phase.

[\[link\]](#) shows how to determine the path length difference for waves traveling from two slits to a common point on a screen. If the screen is a large distance away compared with the distance between the slits, then the angle θ between the path and a line from the slits to the screen (see the figure) is nearly the same for each path. The difference between the paths is shown in the figure; simple trigonometry shows it to be $d \sin \theta$, where d is the distance between the slits. To obtain **constructive interference for a double slit**, the path length difference must be an integral multiple of the wavelength, or

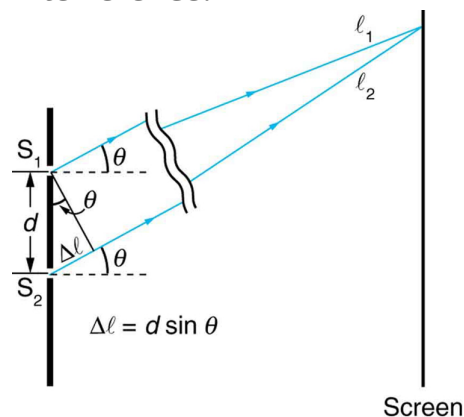
Equation:

$$d \sin \theta = m\lambda, \text{ for } m = 0, 1, -1, 2, -2, \dots \text{ (constructive).}$$

Similarly, to obtain **destructive interference for a double slit**, the path length difference must be a half-integral multiple of the wavelength, or
Equation:

$$d \sin \theta = \left(m + \frac{1}{2}\right) \lambda, \text{ for } m = 0, 1, -1, 2, -2, \dots \text{ (destructive),}$$

where λ is the wavelength of the light, d is the distance between slits, and θ is the angle from the original direction of the beam as discussed above. We call m the **order** of the interference. For example, $m = 4$ is fourth-order interference.



The paths from each slit to a common point on the screen differ by an amount $d \sin \theta$, assuming the distance to the screen is much greater than the distance between slits (not to scale here).

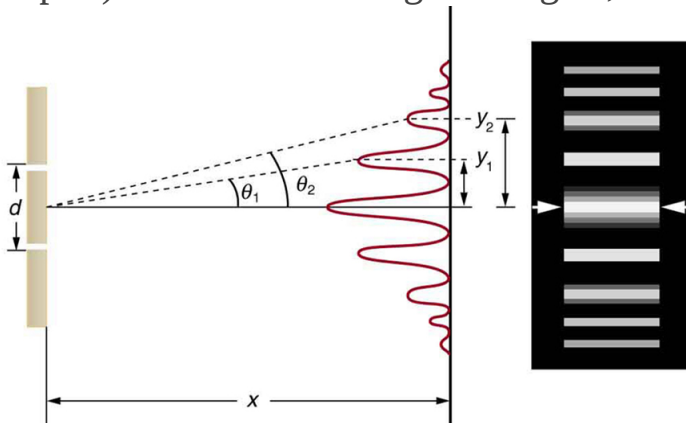
The equations for double slit interference imply that a series of bright and dark lines are formed. For vertical slits, the light spreads out horizontally on

either side of the incident beam into a pattern called interference fringes, illustrated in [\[link\]](#). The intensity of the bright fringes falls off on either side, being brightest at the center. The closer the slits are, the more is the spreading of the bright fringes. We can see this by examining the equation

Equation:

$$d \sin \theta = m\lambda, \text{ for } m = 0, 1, -1, 2, -2, \dots$$

For fixed λ and m , the smaller d is, the larger θ must be, since $\sin \theta = m\lambda / d$. This is consistent with our contention that wave effects are most noticeable when the object the wave encounters (here, slits a distance d apart) is small. Small d gives large θ , hence a large effect.



The interference pattern for a double slit has an intensity that falls off with angle. The photograph shows multiple bright and dark lines, or fringes, formed by light passing through a double slit.

Example:

Finding a Wavelength from an Interference Pattern

Suppose you pass light from a He-Ne laser through two slits separated by 0.0100 mm and find that the third bright line on a screen is formed at an angle of 10.95° relative to the incident beam. What is the wavelength of the light?

Strategy

The third bright line is due to third-order constructive interference, which means that $m = 3$. We are given $d = 0.0100$ mm and $\theta = 10.95^\circ$. The wavelength can thus be found using the equation $d \sin \theta = m\lambda$ for constructive interference.

Solution

The equation is $d \sin \theta = m\lambda$. Solving for the wavelength λ gives

Equation:

$$\lambda = \frac{d \sin \theta}{m}.$$

Substituting known values yields

Equation:

$$\begin{aligned}\lambda &= \frac{(0.0100 \text{ mm})(\sin 10.95^\circ)}{3} \\ &= 6.33 \times 10^{-4} \text{ mm} = 633 \text{ nm}.\end{aligned}$$

Discussion

To three digits, this is the wavelength of light emitted by the common He-Ne laser. Not by coincidence, this red color is similar to that emitted by neon lights. More important, however, is the fact that interference patterns can be used to measure wavelength. Young did this for visible wavelengths. This analytical technique is still widely used to measure electromagnetic spectra. For a given order, the angle for constructive interference increases with λ , so that spectra (measurements of intensity versus wavelength) can be obtained.

Example:

Calculating Highest Order Possible

Interference patterns do not have an infinite number of lines, since there is a limit to how big m can be. What is the highest-order constructive interference possible with the system described in the preceding example?

Strategy and Concept

The equation $d \sin \theta = m\lambda$ (for $m = 0, 1, -1, 2, -2, \dots$) describes constructive interference. For fixed values of d and λ , the larger m is, the larger $\sin \theta$ is. However, the maximum value that $\sin \theta$ can have is 1, for an angle of 90° . (Larger angles imply that light goes backward and does not reach the screen at all.) Let us find which m corresponds to this maximum diffraction angle.

Solution

Solving the equation $d \sin \theta = m\lambda$ for m gives

Equation:

$$m = \frac{d \sin \theta}{\lambda}.$$

Taking $\sin \theta = 1$ and substituting the values of d and λ from the preceding example gives

Equation:

$$m = \frac{(0.0100 \text{ mm})(1)}{633 \text{ nm}} \approx 15.8.$$

Therefore, the largest integer m can be is 15, or

Equation:

$$m = 15.$$

Discussion

The number of fringes depends on the wavelength and slit separation. The number of fringes will be very large for large slit separations. However, if the slit separation becomes much greater than the wavelength, the intensity of the interference pattern changes so that the screen has two bright lines cast by the slits, as expected when light behaves like a ray. We also note that the fringes get fainter further away from the center. Consequently, not all 15 fringes may be observable.

Section Summary

- Young's double slit experiment gave definitive proof of the wave character of light.
- An interference pattern is obtained by the superposition of light from two slits.
- There is constructive interference when $d \sin \theta = m\lambda$ (for $m = 0, 1, -1, 2, -2, \dots$), where d is the distance between the slits, θ is the angle relative to the incident direction, and m is the order of the interference.
- There is destructive interference when $d \sin \theta = (m + \frac{1}{2})\lambda$ (for $m = 0, 1, -1, 2, -2, \dots$).

Conceptual Questions

Exercise:

Problem:

Young's double slit experiment breaks a single light beam into two sources. Would the same pattern be obtained for two independent sources of light, such as the headlights of a distant car? Explain.

Exercise:

Problem:

Suppose you use the same double slit to perform Young's double slit experiment in air and then repeat the experiment in water. Do the angles to the same parts of the interference pattern get larger or smaller? Does the color of the light change? Explain.

Exercise:

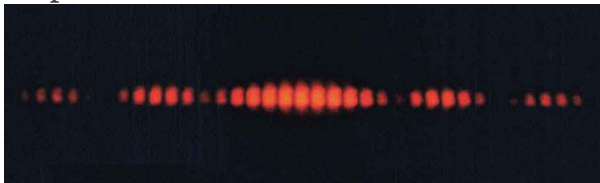
Problem:

Is it possible to create a situation in which there is only destructive interference? Explain.

Exercise:

Problem:

[\[link\]](#) shows the central part of the interference pattern for a pure wavelength of red light projected onto a double slit. The pattern is actually a combination of single slit and double slit interference. Note that the bright spots are evenly spaced. Is this a double slit or single slit characteristic? Note that some of the bright spots are dim on either side of the center. Is this a single slit or double slit characteristic? Which is smaller, the slit width or the separation between slits? Explain your responses.



This double slit interference pattern also shows signs of single slit interference. (credit: PASCO)

Problems & Exercises**Exercise:****Problem:**

At what angle is the first-order maximum for 450-nm wavelength blue light falling on double slits separated by 0.0500 mm?

Solution:

0.516°

Exercise:

Problem:

Calculate the angle for the third-order maximum of 580-nm wavelength yellow light falling on double slits separated by 0.100 mm.

Exercise:**Problem:**

What is the separation between two slits for which 610-nm orange light has its first maximum at an angle of 30.0° ?

Solution:

$$1.22 \times 10^{-6} \text{ m}$$

Exercise:**Problem:**

Find the distance between two slits that produces the first minimum for 410-nm violet light at an angle of 45.0° .

Exercise:**Problem:**

Calculate the wavelength of light that has its third minimum at an angle of 30.0° when falling on double slits separated by $3.00 \mu\text{m}$. Explicitly, show how you follow the steps in [Problem-Solving Strategies for Wave Optics](#).

Solution:

$$600 \text{ nm}$$

Exercise:**Problem:**

What is the wavelength of light falling on double slits separated by $2.00 \mu\text{m}$ if the third-order maximum is at an angle of 60.0° ?

Exercise:

Problem:

At what angle is the fourth-order maximum for the situation in [\[link\]](#)?

Solution:

2.06°

Exercise:

Problem:

What is the highest-order maximum for 400-nm light falling on double slits separated by 25.0 μm ?

Exercise:

Problem:

Find the largest wavelength of light falling on double slits separated by 1.20 μm for which there is a first-order maximum. Is this in the visible part of the spectrum?

Solution:

1200 nm (not visible)

Exercise:

Problem:

What is the smallest separation between two slits that will produce a second-order maximum for 720-nm red light?

Exercise:

Problem:

(a) What is the smallest separation between two slits that will produce a second-order maximum for any visible light? (b) For all visible light?

Solution:

(a) 760 nm

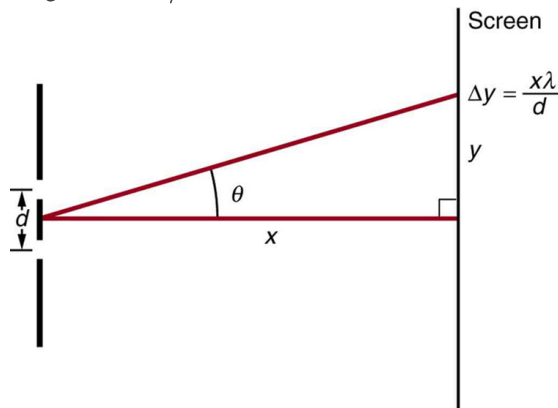
(b) 1520 nm

Exercise:**Problem:**

(a) If the first-order maximum for pure-wavelength light falling on a double slit is at an angle of 10.0° , at what angle is the second-order maximum? (b) What is the angle of the first minimum? (c) What is the highest-order maximum possible here?

Exercise:**Problem:**

[\[link\]](#) shows a double slit located a distance x from a screen, with the distance from the center of the screen given by y . When the distance d between the slits is relatively large, there will be numerous bright spots, called fringes. Show that, for small angles (where $\sin \theta \approx \theta$, with θ in radians), the distance between fringes is given by $\Delta y = x\lambda/d$.



The distance between adjacent fringes is $\Delta y = x\lambda/d$, assuming the

slit separation d is large
compared with λ .

Solution:

For small angles $\sin \theta - \tan \theta \approx \theta$ (in radians).

For two adjacent fringes we have,

Equation:

$$d \sin \theta_m = m\lambda$$

and

Equation:

$$d \sin \theta_{m+1} = (m + 1)\lambda$$

Subtracting these equations gives

Equation:

$$d(\sin \theta_{m+1} - \sin \theta_m) = [(m + 1) - m]\lambda$$

$$d(\theta_{m+1} - \theta_m) = \lambda$$

$$\tan \theta_m = \frac{y_m}{x} \approx \theta_m \Rightarrow d\left(\frac{y_{m+1}}{x} - \frac{y_m}{x}\right) = \lambda$$

$$d \frac{\Delta y}{x} = \lambda \Rightarrow \Delta y = \frac{x\lambda}{d}$$

Exercise:

Problem:

Using the result of the problem above, calculate the distance between fringes for 633-nm light falling on double slits separated by 0.0800 mm, located 3.00 m from a screen as in [\[link\]](#).

Exercise:

Problem:

Using the result of the problem two problems prior, find the wavelength of light that produces fringes 7.50 mm apart on a screen 2.00 m from double slits separated by 0.120 mm (see [\[link\]](#)).

Solution:

450 nm

Glossary

coherent

waves are in phase or have a definite phase relationship

constructive interference for a double slit

the path length difference must be an integral multiple of the wavelength

destructive interference for a double slit

the path length difference must be a half-integral multiple of the wavelength

incoherent

waves have random phase relationships

order

the integer m used in the equations for constructive and destructive interference for a double slit

Introduction to Quantum Physics

class="introduction"

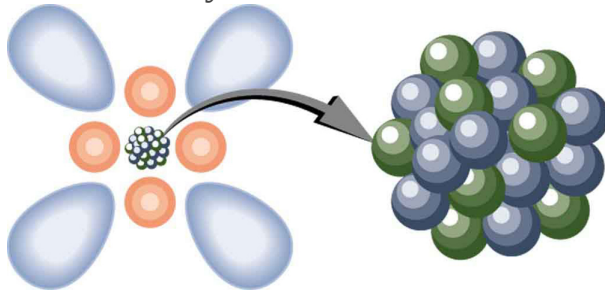
A black fly
imaged by
an electron
microscope
is as
monstrous
as any
science-
fiction
creature.
(credit:
U.S.
Departmen
t of
Agriculture
via
Wikimedia
Commons)



Quantum mechanics is the branch of physics needed to deal with submicroscopic objects. Because these objects are smaller than we can observe directly with our senses and generally must be observed with the aid of instruments, parts of quantum mechanics seem as foreign and bizarre as parts of relativity. But, like relativity, quantum mechanics has been shown to be valid—truth is often stranger than fiction.

Certain aspects of quantum mechanics are familiar to us. We accept as fact that matter is composed of atoms, the smallest unit of an element, and that these atoms combine to form molecules, the smallest unit of a compound. (See [\[link\]](#).) While we cannot see the individual water molecules in a stream, for example, we are aware that this is because molecules are so small and so numerous in that stream. When introducing atoms, we commonly say that electrons orbit atoms in discrete shells around a tiny nucleus, itself composed of smaller particles called protons and neutrons. We are also aware that electric charge comes in tiny units carried almost entirely by electrons and protons. As with water molecules in a stream, we

do not notice individual charges in the current through a lightbulb, because the charges are so small and so numerous in the macroscopic situations we sense directly.



Atoms and their substructure are familiar examples of objects that require quantum mechanics to be fully explained. Certain of their characteristics, such as the discrete electron shells, are classical physics explanations.

In quantum mechanics we conceptualize discrete “electron clouds” around the nucleus.

Note:

Making Connections: Realms of Physics

Classical physics is a good approximation of modern physics under conditions first discussed in the [The Nature of Science and Physics](#).

Quantum mechanics is valid in general, and it must be used rather than classical physics to describe small objects, such as atoms.

Atoms, molecules, and fundamental electron and proton charges are all examples of physical entities that are **quantized**—that is, they appear only in certain discrete values and do not have every conceivable value.

Quantized is the opposite of continuous. We cannot have a fraction of an atom, or part of an electron's charge, or 14-1/3 cents, for example. Rather, everything is built of integral multiples of these substructures. Quantum physics is the branch of physics that deals with small objects and the quantization of various entities, including energy and angular momentum. Just as with classical physics, quantum physics has several subfields, such as mechanics and the study of electromagnetic forces. The **correspondence principle** states that in the classical limit (large, slow-moving objects), **quantum mechanics** becomes the same as classical physics. In this chapter, we begin the development of quantum mechanics and its description of the strange submicroscopic world. In later chapters, we will examine many areas, such as atomic and nuclear physics, in which quantum mechanics is crucial.

Glossary

quantized

the fact that certain physical entities exist only with particular discrete values and not every conceivable value

correspondence principle

in the classical limit (large, slow-moving objects), quantum mechanics becomes the same as classical physics

quantum mechanics

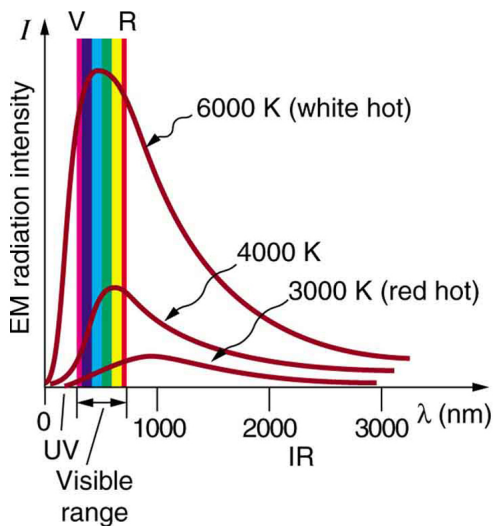
the branch of physics that deals with small objects and with the quantization of various entities, especially energy

Quantization of Energy

- Explain Max Planck's contribution to the development of quantum mechanics.
- Explain why atomic spectra indicate quantization.

Planck's Contribution

Energy is quantized in some systems, meaning that the system can have only certain energies and not a continuum of energies, unlike the classical case. This would be like having only certain speeds at which a car can travel because its kinetic energy can have only certain values. We also find that some forms of energy transfer take place with discrete lumps of energy. While most of us are familiar with the quantization of matter into lumps called atoms, molecules, and the like, we are less aware that energy, too, can be quantized. Some of the earliest clues about the necessity of quantum mechanics over classical physics came from the quantization of energy.



Graphs of blackbody radiation (from an ideal radiator) at three different radiator temperatures. The intensity or rate of

radiation emission
increases dramatically
with temperature, and the
peak of the spectrum
shifts toward the visible
and ultraviolet parts of
the spectrum. The shape
of the spectrum cannot be
described with classical
physics.

Where is the quantization of energy observed? Let us begin by considering the emission and absorption of electromagnetic (EM) radiation. The EM spectrum radiated by a hot solid is linked directly to the solid's temperature. (See [\[link\]](#).) An ideal radiator is one that has an emissivity of 1 at all wavelengths and, thus, is jet black. Ideal radiators are therefore called **blackbodies**, and their EM radiation is called **blackbody radiation**. It was discussed that the total intensity of the radiation varies as T^4 , the fourth power of the absolute temperature of the body, and that the peak of the spectrum shifts to shorter wavelengths at higher temperatures. All of this seems quite continuous, but it was the curve of the spectrum of intensity versus wavelength that gave a clue that the energies of the atoms in the solid are quantized. In fact, providing a theoretical explanation for the experimentally measured shape of the spectrum was a mystery at the turn of the century. When this “ultraviolet catastrophe” was eventually solved, the answers led to new technologies such as computers and the sophisticated imaging techniques described in earlier chapters. Once again, physics as an enabling science changed the way we live.

The German physicist Max Planck (1858–1947) used the idea that atoms and molecules in a body act like oscillators to absorb and emit radiation. The energies of the oscillating atoms and molecules had to be quantized to correctly describe the shape of the blackbody spectrum. Planck deduced that the energy of an oscillator having a frequency f is given by

Equation:

$$E = \left(n + \frac{1}{2} \right) hf.$$

Here n is any nonnegative integer (0, 1, 2, 3, ...). The symbol h stands for **Planck's constant**, given by

Equation:

$$h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}.$$

The equation $E = \left(n + \frac{1}{2} \right) hf$ means that an oscillator having a frequency f (emitting and absorbing EM radiation of frequency f) can have its energy increase or decrease only in *discrete* steps of size

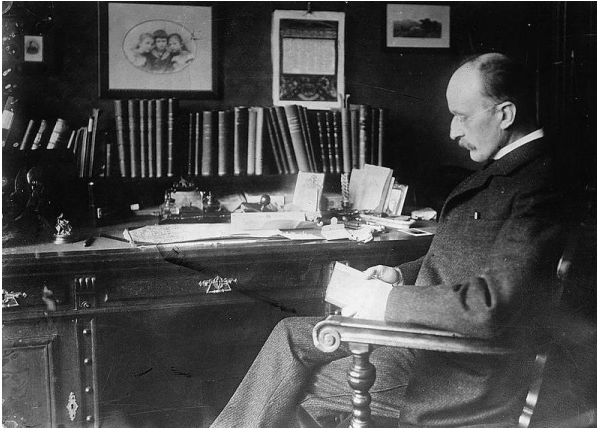
Equation:

$$\Delta E = hf.$$

It might be helpful to mention some macroscopic analogies of this quantization of energy phenomena. This is like a pendulum that has a characteristic oscillation frequency but can swing with only certain amplitudes. Quantization of energy also resembles a standing wave on a string that allows only particular harmonics described by integers. It is also similar to going up and down a hill using discrete stair steps rather than being able to move up and down a continuous slope. Your potential energy takes on discrete values as you move from step to step.

Using the quantization of oscillators, Planck was able to correctly describe the experimentally known shape of the blackbody spectrum. This was the first indication that energy is sometimes quantized on a small scale and earned him the Nobel Prize in Physics in 1918. Although Planck's theory comes from observations of a macroscopic object, its analysis is based on atoms and molecules. It was such a revolutionary departure from classical physics that Planck himself was reluctant to accept his own idea that energy states are not continuous. The general acceptance of Planck's energy quantization was greatly enhanced by Einstein's explanation of the photoelectric effect (discussed in the next section), which took energy

quantization a step further. Planck was fully involved in the development of both early quantum mechanics and relativity. He quickly embraced Einstein's special relativity, published in 1905, and in 1906 Planck was the first to suggest the correct formula for relativistic momentum, $p = \gamma mu$.



The German physicist Max Planck had a major influence on the early development of quantum mechanics, being the first to recognize that energy is sometimes quantized. Planck also made important contributions to special relativity and classical physics.
(credit: Library of Congress, Prints and Photographs Division via Wikimedia Commons)

Note that Planck's constant h is a very small number. So for an infrared frequency of 10^{14} Hz being emitted by a blackbody, for example, the difference between energy levels is only $\Delta E = hf = (6.63 \times 10^{-34} \text{ J}\cdot\text{s})(10^{14} \text{ Hz}) = 6.63 \times 10^{-20} \text{ J}$, or about 0.4 eV. This 0.4 eV of energy is significant compared with typical atomic

energies, which are on the order of an electron volt, or thermal energies, which are typically fractions of an electron volt. But on a macroscopic or classical scale, energies are typically on the order of joules. Even if macroscopic energies are quantized, the quantum steps are too small to be noticed. This is an example of the correspondence principle. For a large object, quantum mechanics produces results indistinguishable from those of classical physics.

Atomic Spectra

Now let us turn our attention to the *emission and absorption of EM radiation by gases*. The Sun is the most common example of a body containing gases emitting an EM spectrum that includes visible light. We also see examples in neon signs and candle flames. Studies of emissions of hot gases began more than two centuries ago, and it was soon recognized that these emission spectra contained huge amounts of information. The type of gas and its temperature, for example, could be determined. We now know that these EM emissions come from electrons transitioning between energy levels in individual atoms and molecules; thus, they are called **atomic spectra**. Atomic spectra remain an important analytical tool today. [\[link\]](#) shows an example of an emission spectrum obtained by passing an electric discharge through a material. One of the most important characteristics of these spectra is that they are discrete. By this we mean that only certain wavelengths, and hence frequencies, are emitted. This is called a line spectrum. If frequency and energy are associated as $\Delta E = hf$, the energies of the electrons in the emitting atoms and molecules are quantized. This is discussed in more detail later in this chapter.



Emission spectrum of oxygen. When an electrical discharge is passed through a substance, its atoms and molecules absorb energy, which is reemitted as EM radiation. The discrete nature of these emissions implies that the energy states of the atoms

and molecules are quantized. Such atomic spectra were used as analytical tools for many decades before it was understood why they are quantized. (credit: Teravolt, Wikimedia Commons)

It was a major puzzle that atomic spectra are quantized. Some of the best minds of 19th-century science failed to explain why this might be. Not until the second decade of the 20th century did an answer based on quantum mechanics begin to emerge. Again a macroscopic or classical body of gas was involved in the studies, but the effect, as we shall see, is due to individual atoms and molecules.

Note:

PhET Explorations: Models of the Hydrogen Atom

How did scientists figure out the structure of atoms without looking at them? Try out different models by shooting light at the atom. Check how the prediction of the model matches the experimental results.

<https://archive.cnx.org/specials/d77cc1d0-33e4-11e6-b016-6726afecd2be/hydrogen-atom/#sim-hydrogen-atom>

Section Summary

- The first indication that energy is sometimes quantized came from blackbody radiation, which is the emission of EM radiation by an object with an emissivity of 1.
- Planck recognized that the energy levels of the emitting atoms and molecules were quantized, with only the allowed values of $E = (n + \frac{1}{2})hf$, where n is any non-negative integer (0, 1, 2, 3, ...).
- h is Planck's constant, whose value is $h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}$.
- Thus, the oscillatory absorption and emission energies of atoms and molecules in a blackbody could increase or decrease only in steps of

size $\Delta E = hf$ where f is the frequency of the oscillatory nature of the absorption and emission of EM radiation.

- Another indication of energy levels being quantized in atoms and molecules comes from the lines in atomic spectra, which are the EM emissions of individual atoms and molecules.

Conceptual Questions

Exercise:

Problem:

Give an example of a physical entity that is quantized. State specifically what the entity is and what the limits are on its values.

Exercise:

Problem:

Give an example of a physical entity that is not quantized, in that it is continuous and may have a continuous range of values.

Exercise:

Problem:

What aspect of the blackbody spectrum forced Planck to propose quantization of energy levels in its atoms and molecules?

Exercise:

Problem:

If Planck's constant were large, say 10^{34} times greater than it is, we would observe macroscopic entities to be quantized. Describe the motions of a child's swing under such circumstances.

Exercise:

Problem: Why don't we notice quantization in everyday events?

Problems & Exercises

Exercise:

Problem:

A LiBr molecule oscillates with a frequency of 1.7×10^{13} Hz. (a) What is the difference in energy in eV between allowed oscillator states? (b) What is the approximate value of n for a state having an energy of 1.0 eV?

Solution:

(a) 0.070 eV

(b) 14

Exercise:

Problem:

The difference in energy between allowed oscillator states in HBr molecules is 0.330 eV. What is the oscillation frequency of this molecule?

Exercise:

Problem:

A physicist is watching a 15-kg orangutan at a zoo swing lazily in a tire at the end of a rope. He (the physicist) notices that each oscillation takes 3.00 s and hypothesizes that the energy is quantized. (a) What is the difference in energy in joules between allowed oscillator states? (b) What is the value of n for a state where the energy is 5.00 J? (c) Can the quantization be observed?

Solution:

(a) 2.21×10^{-34} J

(b) 2.26×10^{34}

(c) No

Glossary

blackbody

an ideal radiator, which can radiate equally well at all wavelengths

blackbody radiation

the electromagnetic radiation from a blackbody

Planck's constant

$$h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}$$

atomic spectra

the electromagnetic emission from atoms and molecules

The Photoelectric Effect

- Describe a typical photoelectric-effect experiment.
- Determine the maximum kinetic energy of photoelectrons ejected by photons of one energy or wavelength, when given the maximum kinetic energy of photoelectrons for a different photon energy or wavelength.

When light strikes materials, it can eject electrons from them. This is called the **photoelectric effect**, meaning that light (*photo*) produces electricity. One common use of the photoelectric effect is in light meters, such as those that adjust the automatic iris on various types of cameras. In a similar way, another use is in solar cells, as you probably have in your calculator or have seen on a roof top or a roadside sign. These make use of the photoelectric effect to convert light into electricity for running different devices.



The
photoelectric
effect can be
observed by
allowing
light to fall
on the metal
plate in this
evacuated
tube.

Electrons
ejected by
the light are
collected on
the collector
wire and

measured as
a current. A
retarding
voltage
between the
collector
wire and
plate can
then be
adjusted so
as to
determine the
energy of the
ejected
electrons. For
example, if it
is sufficiently
negative, no
electrons will
reach the
wire. (credit:
P.P. Urone)

This effect has been known for more than a century and can be studied using a device such as that shown in [\[link\]](#). This figure shows an evacuated tube with a metal plate and a collector wire that are connected by a variable voltage source, with the collector more negative than the plate. When light (or other EM radiation) strikes the plate in the evacuated tube, it may eject electrons. If the electrons have energy in electron volts (eV) greater than the potential difference between the plate and the wire in volts, some electrons will be collected on the wire. Since the electron energy in eV is qV , where q is the electron charge and V is the potential difference, the electron energy can be measured by adjusting the retarding voltage between the wire and the plate. The voltage that stops the electrons from reaching the wire equals the energy in eV. For example, if -3.00 V barely stops the electrons,

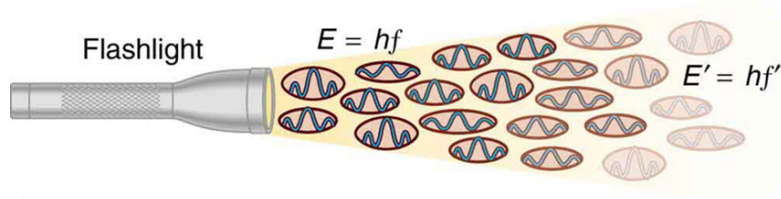
their energy is 3.00 eV. The number of electrons ejected can be determined by measuring the current between the wire and plate. The more light, the more electrons; a little circuitry allows this device to be used as a light meter.

What is really important about the photoelectric effect is what Albert Einstein deduced from it. Einstein realized that there were several characteristics of the photoelectric effect that could be explained only if *EM radiation is itself quantized*: the apparently continuous stream of energy in an EM wave is actually composed of energy quanta called photons. In his explanation of the photoelectric effect, Einstein defined a quantized unit or quantum of EM energy, which we now call a **photon**, with an energy proportional to the frequency of EM radiation. In equation form, the **photon energy** is

Equation:

$$E = hf,$$

where E is the energy of a photon of frequency f and h is Planck's constant. This revolutionary idea looks similar to Planck's quantization of energy states in blackbody oscillators, but it is quite different. It is the quantization of EM radiation itself. EM waves are composed of photons and are not continuous smooth waves as described in previous chapters on optics. Their energy is absorbed and emitted in lumps, not continuously. This is exactly consistent with Planck's quantization of energy levels in blackbody oscillators, since these oscillators increase and decrease their energy in steps of hf by absorbing and emitting photons having $E = hf$. We do not observe this with our eyes, because there are so many photons in common light sources that individual photons go unnoticed. (See [\[link\]](#).) The next section of the text ([Photon Energies and the Electromagnetic Spectrum](#)) is devoted to a discussion of photons and some of their characteristics and implications. For now, we will use the photon concept to explain the photoelectric effect, much as Einstein did.



An EM wave of frequency f is composed of photons, or individual quanta of EM radiation. The energy of each photon is $E = hf$, where h is Planck's constant and f is the frequency of the EM radiation. Higher intensity means more photons per unit area.

The flashlight emits large numbers of photons of many different frequencies, hence others have energy $E' = hf'$, and so on.

The photoelectric effect has the properties discussed below. All these properties are consistent with the idea that individual photons of EM radiation are absorbed by individual electrons in a material, with the electron gaining the photon's energy. Some of these properties are inconsistent with the idea that EM radiation is a simple wave. For simplicity, let us consider what happens with monochromatic EM radiation in which all photons have the same energy hf .

1. If we vary the frequency of the EM radiation falling on a material, we find the following: For a given material, there is a threshold frequency f_0 for the EM radiation below which no electrons are ejected, regardless of intensity. Individual photons interact with individual electrons. Thus if the photon energy is too small to break an electron away, no electrons will be ejected. If EM radiation was a simple wave, sufficient energy could be obtained by increasing the intensity.
2. *Once EM radiation falls on a material, electrons are ejected without delay.* As soon as an individual photon of a sufficiently high frequency is absorbed by an individual electron, the electron is ejected. If the EM

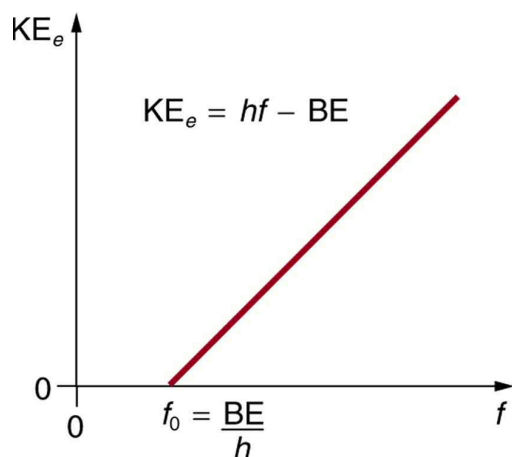
radiation were a simple wave, several minutes would be required for sufficient energy to be deposited to the metal surface to eject an electron.

3. The number of electrons ejected per unit time is proportional to the intensity of the EM radiation and to no other characteristic. High-intensity EM radiation consists of large numbers of photons per unit area, with all photons having the same characteristic energy hf .
4. If we vary the intensity of the EM radiation and measure the energy of ejected electrons, we find the following: *The maximum kinetic energy of ejected electrons is independent of the intensity of the EM radiation.* Since there are so many electrons in a material, it is extremely unlikely that two photons will interact with the same electron at the same time, thereby increasing the energy given it. Instead (as noted in 3 above), increased intensity results in more electrons of the same energy being ejected. If EM radiation were a simple wave, a higher intensity could give more energy, and higher-energy electrons would be ejected.
5. The kinetic energy of an ejected electron equals the photon energy minus the binding energy of the electron in the specific material. An individual photon can give all of its energy to an electron. The photon's energy is partly used to break the electron away from the material. The remainder goes into the ejected electron's kinetic energy. In equation form, this is given by

Equation:

$$KE_e = hf - BE,$$

where KE_e is the maximum kinetic energy of the ejected electron, hf is the photon's energy, and BE is the **binding energy** of the electron to the particular material. (BE is sometimes called the *work function* of the material.) This equation, due to Einstein in 1905, explains the properties of the photoelectric effect quantitatively. An individual photon of EM radiation (it does not come any other way) interacts with an individual electron, supplying enough energy, BE , to break it away, with the remainder going to kinetic energy. The binding energy is $BE = hf_0$, where f_0 is the threshold frequency for the particular material. [\[link\]](#) shows a graph of maximum KE_e versus the frequency of incident EM radiation falling on a particular material.



Photoelectric effect. A graph of the kinetic energy of an ejected electron, KE_e , versus the frequency of EM radiation impinging on a certain material. There is a threshold frequency below which no electrons are ejected, because the individual photon interacting with an individual electron has insufficient energy to break it away. Above the threshold energy, KE_e increases linearly with f , consistent with $KE_e = hf - BE$. The slope of this line is h — the data can be used to determine Planck's constant experimentally. Einstein gave the first successful explanation of such data by proposing

the idea of photons—
quanta of EM radiation.

Einstein's idea that EM radiation is quantized was crucial to the beginnings of quantum mechanics. It is a far more general concept than its explanation of the photoelectric effect might imply. All EM radiation can also be modeled in the form of photons, and the characteristics of EM radiation are entirely consistent with this fact. (As we will see in the next section, many aspects of EM radiation, such as the hazards of ultraviolet (UV) radiation, can be explained *only* by photon properties.) More famous for modern relativity, Einstein planted an important seed for quantum mechanics in 1905, the same year he published his first paper on special relativity. His explanation of the photoelectric effect was the basis for the Nobel Prize awarded to him in 1921. Although his other contributions to theoretical physics were also noted in that award, special and general relativity were not fully recognized in spite of having been partially verified by experiment by 1921. Although hero-worshipped, this great man never received Nobel recognition for his most famous work—relativity.

Example:

Calculating Photon Energy and the Photoelectric Effect: A Violet Light

(a) What is the energy in joules and electron volts of a photon of 420-nm violet light? (b) What is the maximum kinetic energy of electrons ejected from calcium by 420-nm violet light, given that the binding energy (or work function) of electrons for calcium metal is 2.71 eV?

Strategy

To solve part (a), note that the energy of a photon is given by $E = hf$. For part (b), once the energy of the photon is calculated, it is a straightforward application of $KE_e = hf - BE$ to find the ejected electron's maximum kinetic energy, since BE is given.

Solution for (a)

Photon energy is given by

Equation:

$$E = hf$$

Since we are given the wavelength rather than the frequency, we solve the familiar relationship $c = f\lambda$ for the frequency, yielding

Equation:

$$f = \frac{c}{\lambda}.$$

Combining these two equations gives the useful relationship

Equation:

$$E = \frac{hc}{\lambda}.$$

Now substituting known values yields

Equation:

$$E = \frac{(6.63 \times 10^{-34} \text{ J} \cdot \text{s})(3.00 \times 10^8 \text{ m/s})}{420 \times 10^{-9} \text{ m}} = 4.74 \times 10^{-19} \text{ J}.$$

Converting to eV, the energy of the photon is

Equation:

$$E = (4.74 \times 10^{-19} \text{ J}) \frac{1 \text{ eV}}{1.6 \times 10^{-19} \text{ J}} = 2.96 \text{ eV}.$$

Solution for (b)

Finding the kinetic energy of the ejected electron is now a simple application of the equation $\text{KE}_e = hf - \text{BE}$. Substituting the photon energy and binding energy yields

Equation:

$$\text{KE}_e = hf - \text{BE} = 2.96 \text{ eV} - 2.71 \text{ eV} = 0.246 \text{ eV}.$$

Discussion

The energy of this 420-nm photon of violet light is a tiny fraction of a joule, and so it is no wonder that a single photon would be difficult for us to sense directly—humans are more attuned to energies on the order of joules. But looking at the energy in electron volts, we can see that this photon has enough energy to affect atoms and molecules. A DNA molecule can be broken with about 1 eV of energy, for example, and typical atomic and molecular energies are on the order of eV, so that the UV photon in this example could have biological effects. The ejected electron (called a *photoelectron*) has a rather low energy, and it would not travel far, except in a vacuum. The electron would be stopped by a retarding potential of but 0.26 eV. In fact, if the photon wavelength were longer and its energy less than 2.71 eV, then the formula would give a negative kinetic energy, an impossibility. This simply means that the 420-nm photons with their 2.96-eV energy are not much above the frequency threshold. You can show for yourself that the threshold wavelength is 459 nm (blue light). This means that if calcium metal is used in a light meter, the meter will be insensitive to wavelengths longer than those of blue light. Such a light meter would be completely insensitive to red light, for example.

Note:

PhET Explorations: Photoelectric Effect

See how light knocks electrons off a metal target, and recreate the experiment that spawned the field of quantum mechanics.

<https://archive.cnx.org/specials/cf1152da-eae8-11e5-b874-f779884a9994/photoelectric-effect/#sim-photoelectric-effect>

Section Summary

- The photoelectric effect is the process in which EM radiation ejects electrons from a material.
- Einstein proposed photons to be quanta of EM radiation having energy $E = hf$, where f is the frequency of the radiation.

- All EM radiation is composed of photons. As Einstein explained, all characteristics of the photoelectric effect are due to the interaction of individual photons with individual electrons.
- The maximum kinetic energy KE_e of ejected electrons (photoelectrons) is given by $KE_e = hf - BE$, where hf is the photon energy and BE is the binding energy (or work function) of the electron to the particular material.

Conceptual Questions

Exercise:

Problem:

Is visible light the only type of EM radiation that can cause the photoelectric effect?

Exercise:

Problem:

Which aspects of the photoelectric effect cannot be explained without photons? Which can be explained without photons? Are the latter inconsistent with the existence of photons?

Exercise:

Problem:

Is the photoelectric effect a direct consequence of the wave character of EM radiation or of the particle character of EM radiation? Explain briefly.

Exercise:

Problem:

Insulators (nonmetals) have a higher BE than metals, and it is more difficult for photons to eject electrons from insulators. Discuss how this relates to the free charges in metals that make them good conductors.

Exercise:**Problem:**

If you pick up and shake a piece of metal that has electrons in it free to move as a current, no electrons fall out. Yet if you heat the metal, electrons can be boiled off. Explain both of these facts as they relate to the amount and distribution of energy involved with shaking the object as compared with heating it.

Problems & Exercises**Exercise:****Problem:**

What is the longest-wavelength EM radiation that can eject a photoelectron from silver, given that the binding energy is 4.73 eV? Is this in the visible range?

Solution:

263 nm

Exercise:**Problem:**

Find the longest-wavelength photon that can eject an electron from potassium, given that the binding energy is 2.24 eV. Is this visible EM radiation?

Exercise:**Problem:**

What is the binding energy in eV of electrons in magnesium, if the longest-wavelength photon that can eject electrons is 337 nm?

Solution:

3.69 eV

Exercise:

Problem:

Calculate the binding energy in eV of electrons in aluminum, if the longest-wavelength photon that can eject them is 304 nm.

Exercise:

Problem:

What is the maximum kinetic energy in eV of electrons ejected from sodium metal by 450-nm EM radiation, given that the binding energy is 2.28 eV?

Solution:

0.483 eV

Exercise:

Problem:

UV radiation having a wavelength of 120 nm falls on gold metal, to which electrons are bound by 4.82 eV. What is the maximum kinetic energy of the ejected photoelectrons?

Exercise:

Problem:

Violet light of wavelength 400 nm ejects electrons with a maximum kinetic energy of 0.860 eV from sodium metal. What is the binding energy of electrons to sodium metal?

Solution:

2.25 eV

Exercise:

Problem:

UV radiation having a 300-nm wavelength falls on uranium metal, ejecting 0.500-eV electrons. What is the binding energy of electrons to uranium metal?

Exercise:**Problem:**

What is the wavelength of EM radiation that ejects 2.00-eV electrons from calcium metal, given that the binding energy is 2.71 eV? What type of EM radiation is this?

Solution:

- (a) 264 nm
- (b) Ultraviolet

Exercise:**Problem:**

Find the wavelength of photons that eject 0.100-eV electrons from potassium, given that the binding energy is 2.24 eV. Are these photons visible?

Exercise:**Problem:**

What is the maximum velocity of electrons ejected from a material by 80-nm photons, if they are bound to the material by 4.73 eV?

Solution:

$$1.95 \times 10^6 \text{ m/s}$$

Exercise:

Problem:

Photoelectrons from a material with a binding energy of 2.71 eV are ejected by 420-nm photons. Once ejected, how long does it take these electrons to travel 2.50 cm to a detection device?

Exercise:**Problem:**

A laser with a power output of 2.00 mW at a wavelength of 400 nm is projected onto calcium metal. (a) How many electrons per second are ejected? (b) What power is carried away by the electrons, given that the binding energy is 2.71 eV?

Solution:

(a) $4.02 \times 10^{15} \text{ /s}$

(b) 0.256 mW

Exercise:**Problem:**

(a) Calculate the number of photoelectrons per second ejected from a 1.00-mm^2 area of sodium metal by 500-nm EM radiation having an intensity of 1.30 kW/m^2 (the intensity of sunlight above the Earth's atmosphere). (b) Given that the binding energy is 2.28 eV, what power is carried away by the electrons? (c) The electrons carry away less power than brought in by the photons. Where does the other power go? How can it be recovered?

Exercise:**Problem: Unreasonable Results**

Red light having a wavelength of 700 nm is projected onto magnesium metal to which electrons are bound by 3.68 eV. (a) Use $KE_e = hf - BE$ to calculate the kinetic energy of the ejected electrons.

(b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) -1.90 eV

(b) Negative kinetic energy

(c) That the electrons would be knocked free.

Exercise:

Problem: Unreasonable Results

(a) What is the binding energy of electrons to a material from which 4.00-eV electrons are ejected by 400-nm EM radiation? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Glossary

photoelectric effect

the phenomenon whereby some materials eject electrons when light is shined on them

photon

a quantum, or particle, of electromagnetic radiation

photon energy

the amount of energy a photon has; $E = hf$

binding energy

also called the *work function*; the amount of energy necessary to eject an electron from a material

Photon Energies and the Electromagnetic Spectrum

- Explain the relationship between the energy of a photon in joules or electron volts and its wavelength or frequency.
- Calculate the number of photons per second emitted by a monochromatic source of specific wavelength and power.

Ionizing Radiation

A photon is a quantum of EM radiation. Its energy is given by $E = hf$ and is related to the frequency f and wavelength λ of the radiation by

Equation:

$$E = hf = \frac{hc}{\lambda} (\text{energy of a photon}),$$

where E is the energy of a single photon and c is the speed of light. When working with small systems, energy in eV is often useful. Note that Planck's constant in these units is

Equation:

$$h = 4.14 \times 10^{-15} \text{ eV} \cdot \text{s}.$$

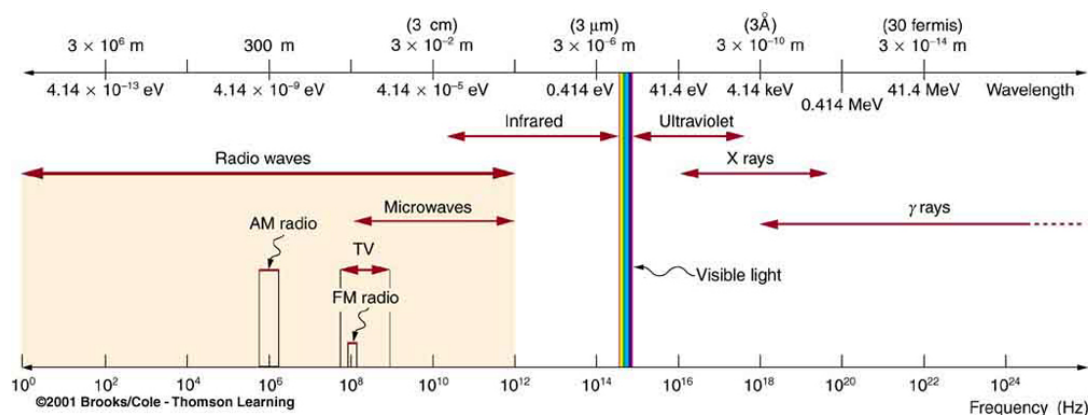
Since many wavelengths are stated in nanometers (nm), it is also useful to know that

Equation:

$$hc = 1240 \text{ eV} \cdot \text{nm}.$$

These will make many calculations a little easier.

All EM radiation is composed of photons. [\[link\]](#) shows various divisions of the EM spectrum plotted against wavelength, frequency, and photon energy. Previously in this book, photon characteristics were alluded to in the discussion of some of the characteristics of UV, x rays, and γ rays, the first of which start with frequencies just above violet in the visible spectrum. It was noted that these types of EM radiation have characteristics much different than visible light. We can now see that such properties arise because photon energy is larger at high frequencies.



The EM spectrum, showing major categories as a function of photon energy in eV, as well as wavelength and frequency. Certain characteristics of EM radiation are directly attributable to photon energy alone.

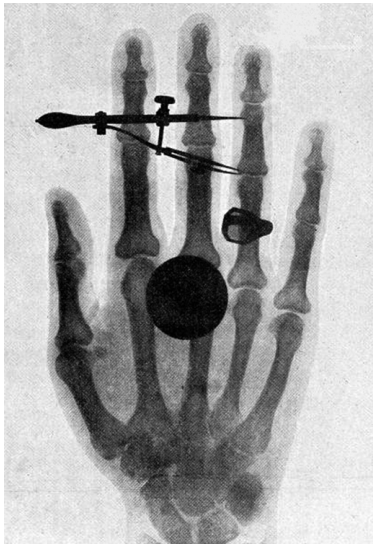
Rotational energies of molecules	10^{-5} eV
Vibrational energies of molecules	0.1 eV
Energy between outer electron shells in atoms	1 eV
Binding energy of a weakly bound molecule	1 eV
Energy of red light	2 eV
Binding energy of a tightly bound molecule	10 eV
Energy to ionize atom or molecule	10 to 1000 eV

Representative Energies for Submicroscopic Effects (Order of Magnitude Only)

Photons act as individual quanta and interact with individual electrons, atoms, molecules, and so on. The energy a photon carries is, thus, crucial to the effects it has. [\[link\]](#) lists representative submicroscopic energies in eV. When we compare photon energies from the EM spectrum in [\[link\]](#) with energies in the table, we can see how effects vary with the type of EM radiation.

Gamma rays, a form of nuclear and cosmic EM radiation, can have the highest frequencies and, hence, the highest photon energies in the EM spectrum. For example, a γ -ray photon with $f = 10^{21}$ Hz has an energy $E = hf = 6.63 \times 10^{-13} \text{ J} = 4.14 \text{ MeV}$. This is sufficient energy to ionize thousands of atoms and molecules, since only 10 to 1000 eV are needed per ionization. In fact, γ rays are one type of **ionizing radiation**, as are x rays and UV, because they produce ionization in materials that absorb

them. Because so much ionization can be produced, a single γ -ray photon can cause significant damage to biological tissue, killing cells or damaging their ability to properly reproduce. When cell reproduction is disrupted, the result can be cancer, one of the known effects of exposure to ionizing radiation. Since cancer cells are rapidly reproducing, they are exceptionally sensitive to the disruption produced by ionizing radiation. This means that ionizing radiation has positive uses in cancer treatment as well as risks in producing cancer.



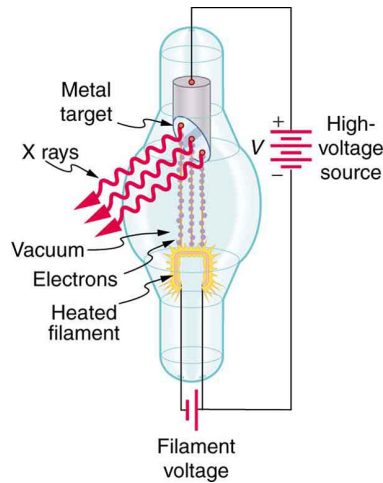
One of the first x-ray images, taken by Röntgen himself. The hand belongs to Bertha Röntgen, his wife.
(credit: Wilhelm Conrad Röntgen, via Wikimedia Commons)

High photon energy also enables γ rays to penetrate materials, since a collision with a single atom or molecule is unlikely to absorb all the γ ray's energy. This can make γ rays useful as a probe, and they are sometimes used in medical imaging. **x rays**, as you can see in [\[link\]](#), overlap with the low-frequency end of the γ ray range. Since x rays have energies of keV and up, individual x-ray photons also can produce large amounts of ionization. At lower photon energies, x rays are not as penetrating as γ rays and are slightly less hazardous. X rays are ideal for medical imaging, their most common use, and a fact that was recognized immediately upon their discovery in 1895 by the German physicist W. C. Roentgen (1845–1923). (See [\[link\]](#).) Within one year of their discovery, x rays (for a time called Roentgen rays) were used for medical diagnostics. Roentgen received the 1901 Nobel Prize for the discovery of x rays.

Note:

Connections: Conservation of Energy

Once again, we find that conservation of energy allows us to consider the initial and final forms that energy takes, without having to make detailed calculations of the intermediate steps. [\[link\]](#) is solved by considering only the initial and final forms of energy.



X rays are produced when energetic electrons strike the copper anode of this cathode ray tube (CRT). Electrons (shown here as separate particles) interact individually with the material they strike, sometimes producing photons of EM radiation.

While γ rays originate in nuclear decay, x rays are produced by the process shown in [\[link\]](#). Electrons ejected by thermal agitation from a hot filament in a vacuum tube are accelerated through a high voltage, gaining kinetic energy from the electrical potential energy. When they strike the anode, the electrons convert their kinetic energy to a variety of forms, including thermal energy. But since an accelerated charge radiates EM waves, and since the electrons act individually, photons are also produced. Some of these x-ray photons obtain the kinetic energy of the electron. The accelerated electrons originate at the cathode, so such a tube is called a cathode ray tube (CRT), and various versions of them are found in older TV and computer screens as well as in x-ray machines.

Example:

X-ray Photon Energy and X-ray Tube Voltage

Find the maximum energy in eV of an x-ray photon produced by electrons accelerated through a potential difference of 50.0 kV in a CRT like the one in [\[link\]](#).

Strategy

Electrons can give all of their kinetic energy to a single photon when they strike the anode of a CRT. (This is something like the photoelectric effect in reverse.) The kinetic energy of the electron comes from electrical potential energy. Thus we can simply equate the maximum photon energy to the electrical potential energy—that is, $hf = qV$. (We do not have to calculate each step from beginning to end if we know that all of the starting energy qV is converted to the final form hf .)

Solution

The maximum photon energy is $hf = qV$, where q is the charge of the electron and V is the accelerating voltage. Thus,

Equation:

$$hf = (1.60 \times 10^{-19} \text{ C})(50.0 \times 10^3 \text{ V}).$$

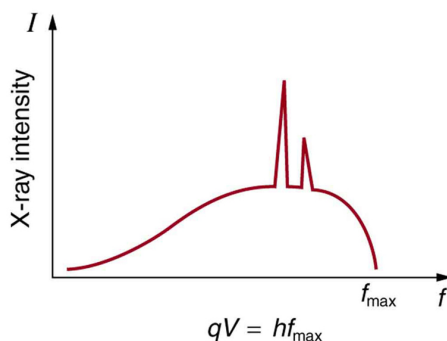
From the definition of the electron volt, we know $1 \text{ eV} = 1.60 \times 10^{-19} \text{ J}$, where $1 \text{ J} = 1 \text{ C} \cdot \text{V}$. Gathering factors and converting energy to eV yields

Equation:

$$hf = (50.0 \times 10^3)(1.60 \times 10^{-19} \text{ C} \cdot \text{V}) \left(\frac{1 \text{ eV}}{1.60 \times 10^{-19} \text{ C} \cdot \text{V}} \right) = (50.0 \times 10^3)(1 \text{ eV}) = 50.0 \text{ keV}.$$

Discussion

This example produces a result that can be applied to many similar situations. If you accelerate a single elementary charge, like that of an electron, through a potential given in volts, then its energy in eV has the same numerical value. Thus a 50.0-kV potential generates 50.0 keV electrons, which in turn can produce photons with a maximum energy of 50 keV. Similarly, a 100-kV potential in an x-ray tube can generate up to 100-keV x-ray photons. Many x-ray tubes have adjustable voltages so that various energy x rays with differing energies, and therefore differing abilities to penetrate, can be generated.



X-ray spectrum obtained when energetic electrons strike a material. The smooth part of the spectrum is bremsstrahlung, while the peaks are characteristic of the anode material. Both are atomic processes that produce energetic

photons known as x-ray photons.

[\[link\]](#) shows the spectrum of x rays obtained from an x-ray tube. There are two distinct features to the spectrum. First, the smooth distribution results from electrons being decelerated in the anode material. A curve like this is obtained by detecting many photons, and it is apparent that the maximum energy is unlikely. This decelerating process produces radiation that is called **bremsstrahlung** (German for *braking radiation*). The second feature is the existence of sharp peaks in the spectrum; these are called **characteristic x rays**, since they are characteristic of the anode material. Characteristic x rays come from atomic excitations unique to a given type of anode material. They are akin to lines in atomic spectra, implying the energy levels of atoms are quantized. Phenomena such as discrete atomic spectra and characteristic x rays are explored further in [Atomic Physics](#).

Ultraviolet radiation (approximately 4 eV to 300 eV) overlaps with the low end of the energy range of x rays, but UV is typically lower in energy. UV comes from the de-excitation of atoms that may be part of a hot solid or gas. These atoms can be given energy that they later release as UV by numerous processes, including electric discharge, nuclear explosion, thermal agitation, and exposure to x rays. A UV photon has sufficient energy to ionize atoms and molecules, which makes its effects different from those of visible light. UV thus has some of the same biological effects as γ rays and x rays. For example, it can cause skin cancer and is used as a sterilizer. The major difference is that several UV photons are required to disrupt cell reproduction or kill a bacterium, whereas single γ -ray and X-ray photons can do the same damage. But since UV does have the energy to alter molecules, it can do what visible light cannot. One of the beneficial aspects of UV is that it triggers the production of vitamin D in the skin, whereas visible light has insufficient energy per photon to alter the molecules that trigger this production. Infantile jaundice is treated by exposing the baby to UV (with eye protection), called phototherapy, the beneficial effects of which are thought to be related to its ability to help prevent the buildup of potentially toxic bilirubin in the blood.

Example:

Photon Energy and Effects for UV

Short-wavelength UV is sometimes called vacuum UV, because it is strongly absorbed by air and must be studied in a vacuum. Calculate the photon energy in eV for 100-nm vacuum UV, and estimate the number of molecules it could ionize or break apart.

Strategy

Using the equation $E = hf$ and appropriate constants, we can find the photon energy and compare it with energy information in [\[link\]](#).

Solution

The energy of a photon is given by

Equation:

$$E = hf = \frac{hc}{\lambda}.$$

Using $hc = 1240 \text{ eV} \cdot \text{nm}$, we find that

Equation:

$$E = \frac{hc}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{100 \text{ nm}} = 12.4 \text{ eV}.$$

Discussion

According to [\[link\]](#), this photon energy might be able to ionize an atom or molecule, and it is about what is needed to break up a tightly bound molecule, since they are bound by approximately 10 eV. This photon energy could destroy about a dozen weakly bound molecules. Because of its high photon energy, UV disrupts atoms and molecules it interacts with. One good consequence is that all but the longest-wavelength UV is strongly absorbed and is easily blocked by sunglasses. In fact, most of the Sun's UV is absorbed by a thin layer of ozone in the upper atmosphere, protecting sensitive organisms on Earth. Damage to our ozone layer by the addition of such chemicals as CFC's has reduced this protection for us.

Visible Light

The range of photon energies for **visible light** from red to violet is 1.63 to 3.26 eV, respectively (left for this chapter's Problems and Exercises to verify). These energies are on the order of those between outer electron shells in atoms and molecules. This means that these photons can be absorbed by atoms and molecules. A *single* photon can actually stimulate the retina, for example, by altering a receptor molecule that then triggers a nerve impulse. Photons can be absorbed or emitted only by atoms and molecules that have precisely the correct quantized energy step to do so. For example, if a red photon of frequency f encounters a molecule that has an energy step, ΔE , equal to hf , then the photon can be absorbed. Violet flowers absorb red and reflect violet; this implies there is no energy step between levels in the receptor molecule equal to the violet photon's energy, but there is an energy step for the red.

There are some noticeable differences in the characteristics of light between the two ends of the visible spectrum that are due to photon energies. Red light has insufficient photon energy to expose most black-and-white film, and it is thus used to illuminate darkrooms where such film is developed. Since violet light has a higher photon energy, dyes that absorb violet tend to fade more quickly than those that do not. (See [\[link\]](#).) Take a look at some faded color posters in a storefront some time, and you will notice that the blues and violets are the last to fade. This is because other dyes, such as red and green dyes, absorb blue and violet photons, the higher energies of which break up their weakly bound molecules. (Complex molecules such as those in dyes and DNA tend to be weakly bound.) Blue and violet dyes reflect those colors and, therefore, do not absorb these more energetic photons, thus suffering less molecular damage.



Why do the reds, yellows,
and greens fade before
the blues and violets
when exposed to the Sun,
as with this poster? The
answer is related to
photon energy. (credit:
Deb Collins, Flickr)

Transparent materials, such as some glasses, do not absorb any visible light, because there is no energy step in the atoms or molecules that could absorb the light. Since individual photons interact with individual atoms, it is nearly impossible to have two photons absorbed simultaneously to reach a large energy step. Because of its lower photon energy, visible light can sometimes pass through many kilometers of a substance, while higher frequencies like UV, x ray, and γ rays are absorbed, because they have sufficient photon energy to ionize the material.

Example:

How Many Photons per Second Does a Typical Light Bulb Produce?

Assuming that 10.0% of a 100-W light bulb's energy output is in the visible range (typical for incandescent bulbs) with an average wavelength of 580 nm, calculate the number of visible photons emitted per second.

Strategy

Power is energy per unit time, and so if we can find the energy per photon, we can determine the number of photons per second. This will best be done in joules, since power is given in watts, which are joules per second.

Solution

The power in visible light production is 10.0% of 100 W, or 10.0 J/s. The energy of the average visible photon is found by substituting the given average wavelength into the formula

Equation:

$$E = \frac{hc}{\lambda}.$$

This produces

Equation:

$$E = \frac{(6.63 \times 10^{-34} \text{ J} \cdot \text{s})(3.00 \times 10^8 \text{ m/s})}{580 \times 10^{-9} \text{ m}} = 3.43 \times 10^{-19} \text{ J}.$$

The number of visible photons per second is thus

Equation:

$$\text{photon/s} = \frac{10.0 \text{ J/s}}{3.43 \times 10^{-19} \text{ J/photon}} = 2.92 \times 10^{19} \text{ photon/s}.$$

Discussion

This incredible number of photons per second is verification that individual photons are insignificant in ordinary human experience. It is also a verification of the correspondence principle—on the macroscopic scale, quantization becomes essentially continuous or classical. Finally, there are so many photons emitted by a 100-W lightbulb that it can be seen by the unaided eye many kilometers away.

Lower-Energy Photons

Infrared radiation (IR) has even lower photon energies than visible light and cannot significantly alter atoms and molecules. IR can be absorbed and emitted by atoms and molecules, particularly between closely spaced states. IR is extremely strongly absorbed by water, for example, because water molecules have many states separated by energies on the order of 10^{-5} eV to 10^{-2} eV, well within the IR and microwave energy ranges. This is why in the IR range, skin is almost jet black, with an emissivity near 1—there are many states in water molecules in the skin that can absorb a large range of IR photon energies. Not all molecules have this property. Air, for example, is nearly transparent to many IR frequencies.

Microwaves are the highest frequencies that can be produced by electronic circuits, although they are also produced naturally. Thus microwaves are similar to IR but do not extend to as high frequencies. There are states in water and other molecules that have the same frequency and energy as microwaves, typically about 10^{-5} eV. This is one reason why food absorbs microwaves more strongly than many other materials, making microwave ovens an efficient way of putting energy directly into food.

Photon energies for both IR and microwaves are so low that huge numbers of photons are involved in any significant energy transfer by IR or microwaves (such as warming yourself with a heat lamp or cooking pizza in the microwave). Visible light, IR, microwaves, and all lower frequencies cannot produce ionization with single photons and do not ordinarily have the hazards of higher frequencies. When visible, IR, or microwave radiation is hazardous, such as the inducement of cataracts by microwaves, the hazard is due to huge numbers of photons acting together (not to an accumulation of photons, such as sterilization by weak UV). The negative effects of visible, IR, or microwave radiation can be thermal effects, which could be produced by any heat source. But one difference is that at very high intensity, strong electric and magnetic fields can be produced by photons acting together. Such electromagnetic fields (EMF) can actually ionize materials.

Note:**Misconception Alert: High-Voltage Power Lines**

Although some people think that living near high-voltage power lines is hazardous to one's health, ongoing studies of the transient field effects produced by these lines show their strengths to be insufficient to cause damage. Demographic studies also fail to show significant correlation of ill effects with high-voltage power lines. The American Physical Society issued a report over 10 years ago on power-line fields, which concluded that the scientific literature and reviews of panels show no consistent, significant link between cancer and power-line fields. They also felt that the "diversion of resources to eliminate a threat which has no persuasive scientific basis is disturbing."

It is virtually impossible to detect individual photons having frequencies below microwave frequencies, because of their low photon energy. But the photons are there. A continuous EM wave can be modeled as photons. At low frequencies, EM waves are generally treated as time- and position-varying electric and magnetic fields with no discernible quantization. This is another example of the correspondence principle in situations involving huge numbers of photons.

Note:**PhET Explorations: Color Vision**

Make a whole rainbow by mixing red, green, and blue light. Change the wavelength of a monochromatic beam or filter white light. View the light as a solid beam, or see the individual photons.

https://phet.colorado.edu/sims/html/color-vision/latest/color-vision_en.html

Section Summary

- Photon energy is responsible for many characteristics of EM radiation, being particularly noticeable at high frequencies.
- Photons have both wave and particle characteristics.

Conceptual Questions**Exercise:**

Problem: Why are UV, x rays, and γ rays called ionizing radiation?

Exercise:**Problem:**

How can treating food with ionizing radiation help keep it from spoiling? UV is not very penetrating. What else could be used?

Exercise:

Problem:

Some television tubes are CRTs. They use an approximately 30-kV accelerating potential to send electrons to the screen, where the electrons stimulate phosphors to emit the light that forms the pictures we watch. Would you expect x rays also to be created?

Exercise:**Problem:**

Tanning salons use “safe” UV with a longer wavelength than some of the UV in sunlight. This “safe” UV has enough photon energy to trigger the tanning mechanism. Is it likely to be able to cause cell damage and induce cancer with prolonged exposure?

Exercise:**Problem:**

Your pupils dilate when visible light intensity is reduced. Does wearing sunglasses that lack UV blockers increase or decrease the UV hazard to your eyes? Explain.

Exercise:**Problem:**

One could feel heat transfer in the form of infrared radiation from a large nuclear bomb detonated in the atmosphere 75 km from you. However, none of the profusely emitted x rays or γ rays reaches you. Explain.

Exercise:

Problem: Can a single microwave photon cause cell damage? Explain.

Exercise:**Problem:**

In an x-ray tube, the maximum photon energy is given by $hf = qV$. Would it be technically more correct to say $hf = qV + BE$, where BE is the binding energy of electrons in the target anode? Why isn't the energy stated the latter way?

Problems & Exercises**Exercise:****Problem:**

What is the energy in joules and eV of a photon in a radio wave from an AM station that has a 1530-kHz broadcast frequency?

Solution:

6.34×10^{-9} eV, 1.01×10^{-27} J

Exercise:

Problem:

(a) Find the energy in joules and eV of photons in radio waves from an FM station that has a 90.0-MHz broadcast frequency. (b) What does this imply about the number of photons per second that the radio station must broadcast?

Exercise:

Problem: Calculate the frequency in hertz of a 1.00-MeV γ -ray photon.

Solution:

$$2.42 \times 10^{20} \text{ Hz}$$

Exercise:**Problem:**

(a) What is the wavelength of a 1.00-eV photon? (b) Find its frequency in hertz. (c) Identify the type of EM radiation.

Exercise:**Problem:**

Do the unit conversions necessary to show that $hc = 1240 \text{ eV} \cdot \text{nm}$, as stated in the text.

Solution:**Equation:**

$$\begin{aligned} hc &= (6.62607 \times 10^{-34} \text{ J} \cdot \text{s}) (2.99792 \times 10^8 \text{ m/s}) \left(\frac{10^9 \text{ nm}}{1 \text{ m}} \right) \left(\frac{1.00000 \text{ eV}}{1.60218 \times 10^{-19} \text{ J}} \right) \\ &= 1239.84 \text{ eV} \cdot \text{nm} \\ &\approx 1240 \text{ eV} \cdot \text{nm} \end{aligned}$$

Exercise:**Problem:**

Confirm the statement in the text that the range of photon energies for visible light is 1.63 to 3.26 eV, given that the range of visible wavelengths is 380 to 760 nm.

Exercise:**Problem:**

(a) Calculate the energy in eV of an IR photon of frequency $2.00 \times 10^{13} \text{ Hz}$. (b) How many of these photons would need to be absorbed simultaneously by a tightly bound molecule to break it apart? (c) What is the energy in eV of a γ ray of frequency $3.00 \times 10^{20} \text{ Hz}$? (d) How many tightly bound molecules could a single such γ ray break apart?

Solution:

(a) 0.0829 eV

- (b) 121
- (c) 1.24 MeV
- (d) 1.24×10^5

Exercise:

Problem: Prove that, to three-digit accuracy, $h = 4.14 \times 10^{-15} \text{ eV} \cdot \text{s}$, as stated in the text.

Exercise:

Problem:

(a) What is the maximum energy in eV of photons produced in a CRT using a 25.0-kV accelerating potential, such as a color TV? (b) What is their frequency?

Solution:

- (a) $25.0 \times 10^3 \text{ eV}$
- (b) $6.04 \times 10^{18} \text{ Hz}$

Exercise:

Problem:

What is the accelerating voltage of an x-ray tube that produces x rays with a shortest wavelength of 0.0103 nm?

Exercise:

Problem:

(a) What is the ratio of power outputs by two microwave ovens having frequencies of 950 and 2560 MHz, if they emit the same number of photons per second? (b) What is the ratio of photons per second if they have the same power output?

Solution:

- (a) 2.69
- (b) 0.371

Exercise:

Problem:

How many photons per second are emitted by the antenna of a microwave oven, if its power output is 1.00 kW at a frequency of 2560 MHz?

Exercise:

Problem:

Some satellites use nuclear power. (a) If such a satellite emits a 1.00-W flux of γ rays having an average energy of 0.500 MeV, how many are emitted per second? (b) These γ rays affect other satellites. How far away must another satellite be to only receive one γ ray per second per square meter?

Solution:

(a) 1.25×10^{13} photons/s

(b) 997 km

Exercise:**Problem:**

(a) If the power output of a 650-kHz radio station is 50.0 kW, how many photons per second are produced? (b) If the radio waves are broadcast uniformly in all directions, find the number of photons per second per square meter at a distance of 100 km. Assume no reflection from the ground or absorption by the air.

Exercise:**Problem:**

How many x-ray photons per second are created by an x-ray tube that produces a flux of x rays having a power of 1.00 W? Assume the average energy per photon is 75.0 keV.

Solution:

8.33×10^{13} photons/s

Exercise:**Problem:**

(a) How far away must you be from a 650-kHz radio station with power 50.0 kW for there to be only one photon per second per square meter? Assume no reflections or absorption, as if you were in deep outer space. (b) Discuss the implications for detecting intelligent life in other solar systems by detecting their radio broadcasts.

Exercise:**Problem:**

Assuming that 10.0% of a 100-W light bulb's energy output is in the visible range (typical for incandescent bulbs) with an average wavelength of 580 nm, and that the photons spread out uniformly and are not absorbed by the atmosphere, how far away would you be if 500 photons per second enter the 3.00-mm diameter pupil of your eye? (This number easily stimulates the retina.)

Solution:

181 km

Exercise:

Problem:Construct Your Own Problem

Consider a laser pen. Construct a problem in which you calculate the number of photons per second emitted by the pen. Among the things to be considered are the laser pen's wavelength and power output. Your instructor may also wish for you to determine the minimum diffraction spreading in the beam and the number of photons per square centimeter the pen can project at some large distance. In this latter case, you will also need to consider the output size of the laser beam, the distance to the object being illuminated, and any absorption or scattering along the way.

Glossary

gamma ray

also γ -ray; highest-energy photon in the EM spectrum

ionizing radiation

radiation that ionizes materials that absorb it

x ray

EM photon between γ -ray and UV in energy

bremsstrahlung

German for *braking radiation*; produced when electrons are decelerated

characteristic x rays

x rays whose energy depends on the material they were produced in

ultraviolet radiation

UV; ionizing photons slightly more energetic than violet light

visible light

the range of photon energies the human eye can detect

infrared radiation

photons with energies slightly less than red light

microwaves

photons with wavelengths on the order of a micron (μm)

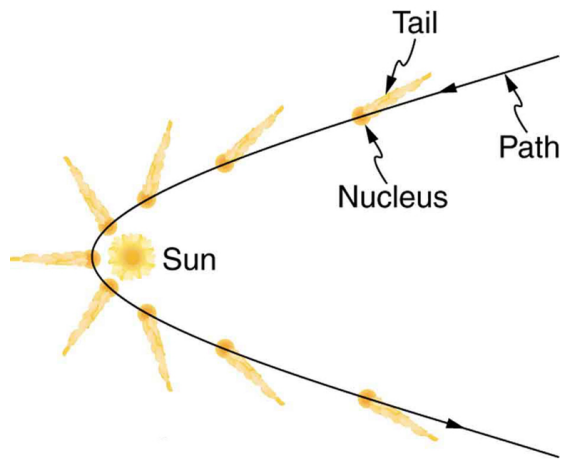
Photon Momentum

- Relate the linear momentum of a photon to its energy or wavelength, and apply linear momentum conservation to simple processes involving the emission, absorption, or reflection of photons.
- Account qualitatively for the increase of photon wavelength that is observed, and explain the significance of the Compton wavelength.

Measuring Photon Momentum

The quantum of EM radiation we call a **photon** has properties analogous to those of particles we can see, such as grains of sand. A photon interacts as a unit in collisions or when absorbed, rather than as an extensive wave.

Massive quanta, like electrons, also act like macroscopic particles—something we expect, because they are the smallest units of matter. Particles carry momentum as well as energy. Despite photons having no mass, there has long been evidence that EM radiation carries momentum. (Maxwell and others who studied EM waves predicted that they would carry momentum.) It is now a well-established fact that photons *do* have momentum. In fact, photon momentum is suggested by the photoelectric effect, where photons knock electrons out of a substance. [\[link\]](#) shows macroscopic evidence of photon momentum.



The tails of the Hale-Bopp comet point away from the Sun, evidence that light has momentum. Dust emanating from the body of the comet forms this tail. Particles of dust are pushed away from the Sun by light reflecting from them. The blue ionized gas tail is also produced by photons interacting with atoms in the comet material. (credit: Geoff Chester, U.S. Navy, via Wikimedia Commons)

[\[link\]](#) shows a comet with two prominent tails. What most people do not know about the tails is that they always point *away* from the Sun rather than trailing behind the comet (like the tail of Bo Peep's sheep). Comet tails are composed of gases and dust evaporated from the body of the comet and ionized gas. The dust particles recoil away from the Sun when photons scatter from them. Evidently, photons carry momentum in the direction of their motion (away from the Sun), and some of this momentum is transferred to dust particles in collisions. Gas atoms and molecules in the blue tail are most affected by other particles of radiation, such as protons and electrons emanating from the Sun, rather than by the momentum of photons.

Note:

Connections: Conservation of Momentum

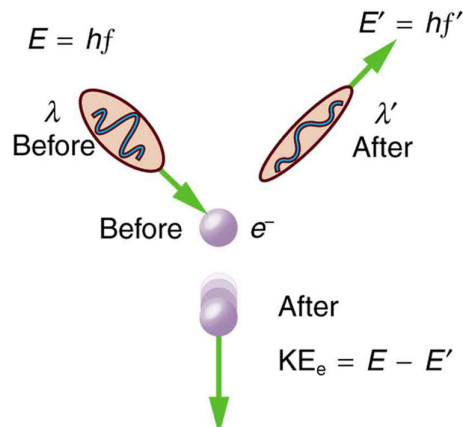
Not only is momentum conserved in all realms of physics, but all types of particles are found to have momentum. We expect particles with mass to have momentum, but now we see that massless particles including photons also carry momentum.

Momentum is conserved in quantum mechanics just as it is in relativity and classical physics. Some of the earliest direct experimental evidence of this came from scattering of x-ray photons by electrons in substances, named Compton scattering after the American physicist Arthur H. Compton (1892–1962). Around 1923, Compton observed that x rays scattered from materials had a decreased energy and correctly analyzed this as being due to the scattering of photons from electrons. This phenomenon could be handled as a collision between two particles—a photon and an electron at rest in the material. Energy and momentum are conserved in the collision. (See [\[link\]](#)) He won a Nobel Prize in 1929 for the discovery of this scattering, now called the **Compton effect**, because it helped prove that **photon momentum** is given by

Equation:

$$p = \frac{h}{\lambda},$$

where h is Planck's constant and λ is the photon wavelength. (Note that relativistic momentum given as $p = \gamma mu$ is valid only for particles having mass.)



The Compton effect is the name given to the scattering of a photon by an electron. Energy and momentum are conserved, resulting in a reduction of both for the scattered photon. Studying this effect, Compton verified that photons have momentum.

We can see that photon momentum is small, since $p = h/\lambda$ and h is very small. It is for this reason that we do not ordinarily observe photon

momentum. Our mirrors do not recoil when light reflects from them (except perhaps in cartoons). Compton saw the effects of photon momentum because he was observing x rays, which have a small wavelength and a relatively large momentum, interacting with the lightest of particles, the electron.

Example:

Electron and Photon Momentum Compared

(a) Calculate the momentum of a visible photon that has a wavelength of 500 nm. (b) Find the velocity of an electron having the same momentum. (c) What is the energy of the electron, and how does it compare with the energy of the photon?

Strategy

Finding the photon momentum is a straightforward application of its definition: $p = \frac{h}{\lambda}$. If we find the photon momentum is small, then we can assume that an electron with the same momentum will be nonrelativistic, making it easy to find its velocity and kinetic energy from the classical formulas.

Solution for (a)

Photon momentum is given by the equation:

Equation:

$$p = \frac{h}{\lambda}.$$

Entering the given photon wavelength yields

Equation:

$$p = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{500 \times 10^{-9} \text{ m}} = 1.33 \times 10^{-27} \text{ kg} \cdot \text{m/s}.$$

Solution for (b)

Since this momentum is indeed small, we will use the classical expression $p = mv$ to find the velocity of an electron with this momentum. Solving for v and using the known value for the mass of an electron gives

Equation:

$$v = \frac{p}{m} = \frac{1.33 \times 10^{-27} \text{ kg} \cdot \text{m/s}}{9.11 \times 10^{-31} \text{ kg}} = 1460 \text{ m/s} \approx 1460 \text{ m/s}.$$

Solution for (c)

The electron has kinetic energy, which is classically given by

Equation:

$$\text{KE}_e = \frac{1}{2}mv^2.$$

Thus,

Equation:

$$\text{KE}_e = \frac{1}{2}(9.11 \times 10^{-31} \text{ kg})(1455 \text{ m/s})^2 = 9.64 \times 10^{-25} \text{ J}.$$

Converting this to eV by multiplying by $(1 \text{ eV})/(1.602 \times 10^{-19} \text{ J})$ yields

Equation:

$$\text{KE}_e = 6.02 \times 10^{-6} \text{ eV}.$$

The photon energy E is

Equation:

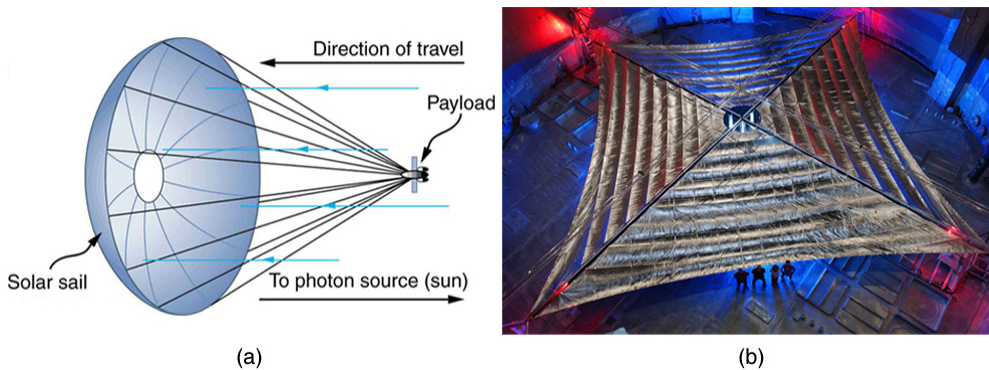
$$E = \frac{hc}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{500 \text{ nm}} = 2.48 \text{ eV},$$

which is about five orders of magnitude greater.

Discussion

Photon momentum is indeed small. Even if we have huge numbers of them, the total momentum they carry is small. An electron with the same momentum has a 1460 m/s velocity, which is clearly nonrelativistic. A more massive particle with the same momentum would have an even smaller velocity. This is borne out by the fact that it takes far less energy to give an electron the same momentum as a photon. But on a quantum-mechanical scale, especially for high-energy photons interacting with small

masses, photon momentum is significant. Even on a large scale, photon momentum can have an effect if there are enough of them and if there is nothing to prevent the slow recoil of matter. Comet tails are one example, but there are also proposals to build space sails that use huge low-mass mirrors (made of aluminized Mylar) to reflect sunlight. In the vacuum of space, the mirrors would gradually recoil and could actually take spacecraft from place to place in the solar system. (See [\[link\]](#).)



(a) Space sails have been proposed that use the momentum of sunlight reflecting from gigantic low-mass sails to propel spacecraft about the solar system. A Russian test model of this (the Cosmos 1) was launched in 2005, but did not make it into orbit due to a rocket failure. (b) A U.S. version of this, labeled LightSail-1, is scheduled for trial launches in the first part of this decade. It will have a 40-m² sail. (credit: Kim Newton/NASA)

Relativistic Photon Momentum

There is a relationship between photon momentum p and photon energy E that is consistent with the relation given previously for the relativistic total energy of a particle as $E^2 = (pc)^2 + (mc)^2$. We know m is zero for a photon, but p is not, so that $E^2 = (pc)^2 + (mc)^2$ becomes

Equation:

$$E = pc,$$

or

Equation:

$$p = \frac{E}{c} \text{ (photons).}$$

To check the validity of this relation, note that $E = hc/\lambda$ for a photon. Substituting this into $p = E/c$ yields

Equation:

$$p = (hc/\lambda)/c = \frac{h}{\lambda},$$

as determined experimentally and discussed above. Thus, $p = E/c$ is equivalent to Compton's result $p = h/\lambda$. For a further verification of the relationship between photon energy and momentum, see [\[link\]](#).

Note:

Photon Detectors

Almost all detection systems talked about thus far—eyes, photographic plates, photomultiplier tubes in microscopes, and CCD cameras—rely on particle-like properties of photons interacting with a sensitive area. A change is caused and either the change is cascaded or zillions of points are recorded to form an image we detect. These detectors are used in biomedical imaging systems, and there is ongoing research into improving the efficiency of receiving photons, particularly by cooling detection systems and reducing thermal effects.

Example:**Photon Energy and Momentum**

Show that $p = E/c$ for the photon considered in the [\[link\]](#).

Strategy

We will take the energy E found in [\[link\]](#), divide it by the speed of light, and see if the same momentum is obtained as before.

Solution

Given that the energy of the photon is 2.48 eV and converting this to joules, we get

Equation:

$$p = \frac{E}{c} = \frac{(2.48 \text{ eV})(1.60 \times 10^{-19} \text{ J/eV})}{3.00 \times 10^8 \text{ m/s}} = 1.33 \times 10^{-27} \text{ kg} \cdot \text{m/s}.$$

Discussion

This value for momentum is the same as found before (note that unrounded values are used in all calculations to avoid even small rounding errors), an expected verification of the relationship $p = E/c$. This also means the relationship between energy, momentum, and mass given by $E^2 = (pc)^2 + (mc)^2$ applies to both matter and photons. Once again, note that p is not zero, even when m is.

Note:**Problem-Solving Suggestion**

Note that the forms of the constants $h = 4.14 \times 10^{-15} \text{ eV} \cdot \text{s}$ and $hc = 1240 \text{ eV} \cdot \text{nm}$ may be particularly useful for this section's Problems and Exercises.

Section Summary

- Photons have momentum, given by $p = \frac{h}{\lambda}$, where λ is the photon wavelength.

- Photon energy and momentum are related by $p = \frac{E}{c}$, where $E = hf = hc/\lambda$ for a photon.

Conceptual Questions

Exercise:

Problem:

Which formula may be used for the momentum of all particles, with or without mass?

Exercise:

Problem:

Is there any measurable difference between the momentum of a photon and the momentum of matter?

Exercise:

Problem:

Why don't we feel the momentum of sunlight when we are on the beach?

Problems & Exercises

Exercise:

Problem:

- (a) Find the momentum of a 4.00-cm-wavelength microwave photon.
- (b) Discuss why you expect the answer to (a) to be very small.

Solution:

- (a) $1.66 \times 10^{-32} \text{ kg} \cdot \text{m/s}$

(b) The wavelength of microwave photons is large, so the momentum they carry is very small.

Exercise:

Problem:

(a) What is the momentum of a 0.0100-nm-wavelength photon that could detect details of an atom? (b) What is its energy in MeV?

Exercise:

Problem:

(a) What is the wavelength of a photon that has a momentum of $5.00 \times 10^{-29} \text{ kg} \cdot \text{m/s}$? (b) Find its energy in eV.

Solution:

(a) $13.3 \text{ } \mu\text{m}$

(b) $9.38 \times 10^{-2} \text{ eV}$

Exercise:

Problem:

(a) A γ -ray photon has a momentum of $8.00 \times 10^{-21} \text{ kg} \cdot \text{m/s}$. What is its wavelength? (b) Calculate its energy in MeV.

Exercise:

Problem:

(a) Calculate the momentum of a photon having a wavelength of $2.50 \text{ } \mu\text{m}$. (b) Find the velocity of an electron having the same momentum. (c) What is the kinetic energy of the electron, and how does it compare with that of the photon?

Solution:

(a) $2.65 \times 10^{-28} \text{ kg} \cdot \text{m/s}$

(b) 291 m/s

(c) electron 3.86×10^{-26} J, photon 7.96×10^{-20} J, ratio 2.06×10^6

Exercise:

Problem:

Repeat the previous problem for a 10.0-nm-wavelength photon.

Exercise:

Problem:

(a) Calculate the wavelength of a photon that has the same momentum as a proton moving at 1.00% of the speed of light. (b) What is the energy of the photon in MeV? (c) What is the kinetic energy of the proton in MeV?

Solution:

(a) 1.32×10^{-13} m

(b) 9.39 MeV

(c) 4.70×10^{-2} MeV

Exercise:

Problem:

(a) Find the momentum of a 100-keV x-ray photon. (b) Find the equivalent velocity of a neutron with the same momentum. (c) What is the neutron's kinetic energy in keV?

Exercise:

Problem:

Take the ratio of relativistic rest energy, $E = \gamma mc^2$, to relativistic momentum, $p = \gamma mu$, and show that in the limit that mass approaches zero, you find $E/p = c$.

Solution:

$E = \gamma mc^2$ and $P = \gamma mu$, so

Equation:

$$\frac{E}{P} = \frac{\gamma mc^2}{\gamma mu} = \frac{c^2}{u}.$$

As the mass of particle approaches zero, its velocity u will approach c , so that the ratio of energy to momentum in this limit is

Equation:

$$\lim_{m \rightarrow 0} \frac{E}{P} = \frac{c^2}{c} = c$$

which is consistent with the equation for photon energy.

Exercise:**Problem: Construct Your Own Problem**

Consider a space sail such as mentioned in [\[link\]](#). Construct a problem in which you calculate the light pressure on the sail in N/m^2 produced by reflecting sunlight. Also calculate the force that could be produced and how much effect that would have on a spacecraft. Among the things to be considered are the intensity of sunlight, its average wavelength, the number of photons per square meter this implies, the area of the space sail, and the mass of the system being accelerated.

Exercise:**Problem: Unreasonable Results**

A car feels a small force due to the light it sends out from its headlights, equal to the momentum of the light divided by the time in which it is emitted. (a) Calculate the power of each headlight, if they

exert a total force of 2.00×10^{-2} N backward on the car. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) 3.00×10^6 W

(b) Headlights are way too bright.

(c) Force is too large.

Glossary

photon momentum

the amount of momentum a photon has, calculated by $p = \frac{h}{\lambda} = \frac{E}{c}$

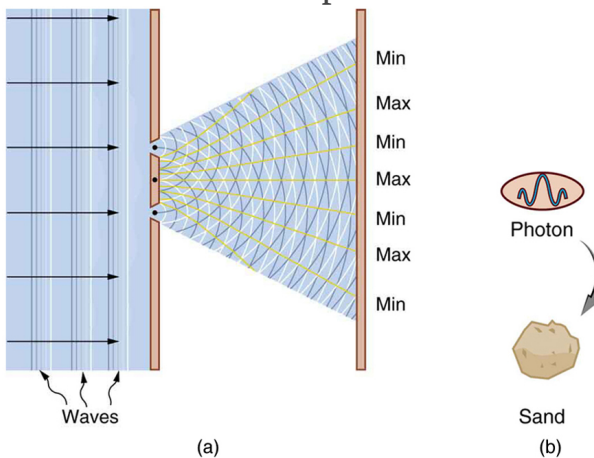
Compton effect

the phenomenon whereby x rays scattered from materials have decreased energy

The Particle-Wave Duality

- Explain what the term particle-wave duality means, and why it is applied to EM radiation.

We have long known that EM radiation is a wave, capable of interference and diffraction. We now see that light can be modeled as photons, which are massless particles. This may seem contradictory, since we ordinarily deal with large objects that never act like both wave and particle. An ocean wave, for example, looks nothing like a rock. To understand small-scale phenomena, we make analogies with the large-scale phenomena we observe directly. When we say something behaves like a wave, we mean it shows interference effects analogous to those seen in overlapping water waves. (See [\[link\]](#).) Two examples of waves are sound and EM radiation. When we say something behaves like a particle, we mean that it interacts as a discrete unit with no interference effects. Examples of particles include electrons, atoms, and photons of EM radiation. How do we talk about a phenomenon that acts like both a particle and a wave?



(a) The interference pattern for light through a double slit is a wave property understood by analogy to water waves. (b) The properties of photons having quantized energy and momentum and acting as a concentrated unit are

understood by analogy to
macroscopic particles.

There is no doubt that EM radiation interferes and has the properties of wavelength and frequency. There is also no doubt that it behaves as particles—photons with discrete energy. We call this twofold nature the **particle-wave duality**, meaning that EM radiation has both particle and wave properties. This so-called duality is simply a term for properties of the photon analogous to phenomena we can observe directly, on a macroscopic scale. If this term seems strange, it is because we do not ordinarily observe details on the quantum level directly, and our observations yield either particle *or* wavelike properties, but never both simultaneously.

Since we have a particle-wave duality for photons, and since we have seen connections between photons and matter in that both have momentum, it is reasonable to ask whether there is a particle-wave duality for matter as well. If the EM radiation we once thought to be a pure wave has particle properties, is it possible that matter has wave properties? The answer is yes. The consequences are tremendous, as we will begin to see in the next section.

Note:

PhET Explorations: Quantum Wave Interference

When do photons, electrons, and atoms behave like particles and when do they behave like waves? Watch waves spread out and interfere as they pass through a double slit, then get detected on a screen as tiny dots. Use quantum detectors to explore how measurements change the waves and the patterns they produce on the screen.

[Quantum](#)
[Wave](#)
[Interferenc](#)
[e](#)



Section Summary

- EM radiation can behave like either a particle or a wave.
- This is termed particle-wave duality.

Glossary

particle-wave duality

the property of behaving like either a particle or a wave; the term for the phenomenon that all particles have wave characteristics

The Wave Nature of Matter

- Describe the Davisson-Germer experiment, and explain how it provides evidence for the wave nature of electrons.

De Broglie Wavelength

In 1923 a French physics graduate student named Prince Louis-Victor de Broglie (1892–1987) made a radical proposal based on the hope that nature is symmetric. If EM radiation has both particle and wave properties, then nature would be symmetric if matter also had both particle and wave properties. If what we once thought of as an unequivocal wave (EM radiation) is also a particle, then what we think of as an unequivocal particle (matter) may also be a wave. De Broglie's suggestion, made as part of his doctoral thesis, was so radical that it was greeted with some skepticism. A copy of his thesis was sent to Einstein, who said it was not only probably correct, but that it might be of fundamental importance. With the support of Einstein and a few other prominent physicists, de Broglie was awarded his doctorate.

De Broglie took both relativity and quantum mechanics into account to develop the proposal that *all particles have a wavelength*, given by

Equation:

$$\lambda = \frac{h}{p} \text{ (matter and photons),}$$

where h is Planck's constant and p is momentum. This is defined to be the **de Broglie wavelength**. (Note that we already have this for photons, from the equation $p = h/\lambda$.) The hallmark of a wave is interference. If matter is a wave, then it must exhibit constructive and destructive interference. Why isn't this ordinarily observed? The answer is that in order to see significant interference effects, a wave must interact with an object about the same size as its wavelength. Since h is very small, λ is also small, especially for macroscopic objects. A 3-kg bowling ball moving at 10 m/s, for example, has

Equation:

$$\lambda = h/p = (6.63 \times 10^{-34} \text{ J}\cdot\text{s})/[(3 \text{ kg})(10 \text{ m/s})] = 2 \times 10^{-35} \text{ m}.$$

This means that to see its wave characteristics, the bowling ball would have to interact with something about 10^{-35} m in size—far smaller than anything known. When waves interact with objects much larger than their wavelength, they show negligible interference effects and move in straight lines (such as light rays in geometric optics). To get easily observed interference effects from particles of matter, the longest wavelength and hence smallest mass possible would be useful. Therefore, this effect was first observed with electrons.

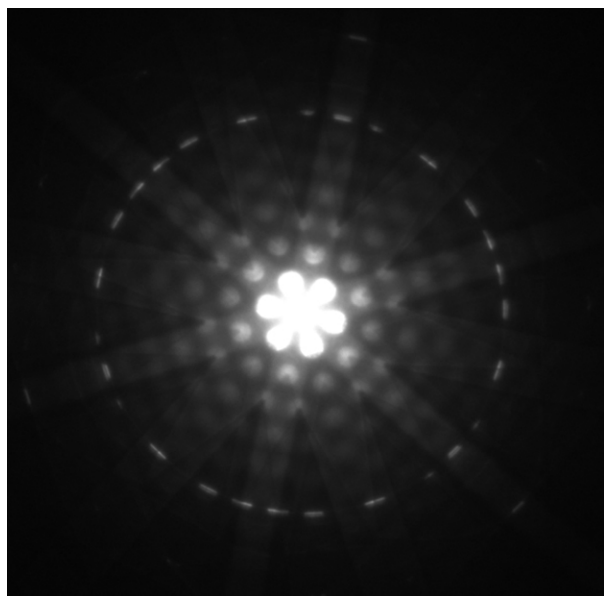
American physicists Clinton J. Davisson and Lester H. Germer in 1925 and, independently, British physicist G. P. Thomson (son of J. J. Thomson, discoverer of the electron) in 1926 scattered electrons from crystals and found diffraction patterns. These patterns are exactly consistent with interference of electrons having the de Broglie wavelength and are somewhat analogous to light interacting with a diffraction grating. (See [\[link\]](#).)

Note:**Connections: Waves**

All microscopic particles, whether massless, like photons, or having mass, like electrons, have wave properties. The relationship between momentum and wavelength is fundamental for all particles.

De Broglie's proposal of a wave nature for all particles initiated a remarkably productive era in which the foundations for quantum mechanics were laid. In 1926, the Austrian physicist Erwin Schrödinger (1887–1961) published four papers in which the wave nature of particles was treated explicitly with wave equations. At the same time, many others began important work. Among them was German physicist Werner Heisenberg

(1901–1976) who, among many other contributions to quantum mechanics, formulated a mathematical treatment of the wave nature of matter that used matrices rather than wave equations. We will deal with some specifics in later sections, but it is worth noting that de Broglie's work was a watershed for the development of quantum mechanics. De Broglie was awarded the Nobel Prize in 1929 for his vision, as were Davisson and G. P. Thomson in 1937 for their experimental verification of de Broglie's hypothesis.



This diffraction pattern was obtained for electrons diffracted by crystalline silicon. Bright regions are those of constructive interference, while dark regions are those of destructive interference. (credit: Ndtthe, Wikimedia Commons)

Example:

Electron Wavelength versus Velocity and Energy

For an electron having a de Broglie wavelength of 0.167 nm (appropriate for interacting with crystal lattice structures that are about this size): (a)

Calculate the electron's velocity, assuming it is nonrelativistic. (b)

Calculate the electron's kinetic energy in eV.

Strategy

For part (a), since the de Broglie wavelength is given, the electron's velocity can be obtained from $\lambda = h/p$ by using the nonrelativistic formula for momentum, $p = mv$. For part (b), once v is obtained (and it has been verified that v is nonrelativistic), the classical kinetic energy is simply $(1/2)mv^2$.

Solution for (a)

Substituting the nonrelativistic formula for momentum ($p = mv$) into the de Broglie wavelength gives

Equation:

$$\lambda = \frac{h}{p} = \frac{h}{mv}.$$

Solving for v gives

Equation:

$$v = \frac{h}{m\lambda}.$$

Substituting known values yields

Equation:

$$v = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{(9.11 \times 10^{-31} \text{ kg})(0.167 \times 10^{-9} \text{ m})} = 4.36 \times 10^6 \text{ m/s}.$$

Solution for (b)

While fast compared with a car, this electron's speed is not highly relativistic, and so we can comfortably use the classical formula to find the electron's kinetic energy and convert it to eV as requested.

Equation:

$$\begin{aligned}
 \text{KE} &= \frac{1}{2}mv^2 \\
 &= \frac{1}{2}(9.11 \times 10^{-31} \text{ kg})(4.36 \times 10^6 \text{ m/s})^2 \\
 &= (86.4 \times 10^{-18} \text{ J}) \left(\frac{1 \text{ eV}}{1.602 \times 10^{-19} \text{ J}} \right) \\
 &= 54.0 \text{ eV}
 \end{aligned}$$

Discussion

This low energy means that these 0.167-nm electrons could be obtained by accelerating them through a 54.0-V electrostatic potential, an easy task. The results also confirm the assumption that the electrons are nonrelativistic, since their velocity is just over 1% of the speed of light and the kinetic energy is about 0.01% of the rest energy of an electron (0.511 MeV). If the electrons had turned out to be relativistic, we would have had to use more involved calculations employing relativistic formulas.

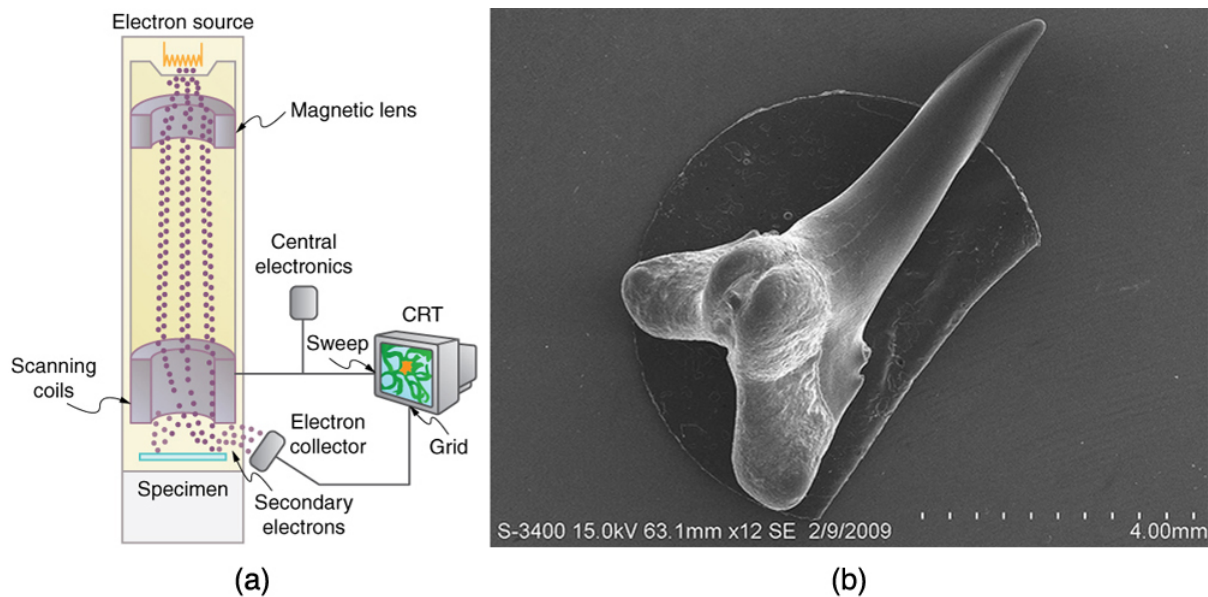
Electron Microscopes

One consequence or use of the wave nature of matter is found in the electron microscope. As we have discussed, there is a limit to the detail observed with any probe having a wavelength. Resolution, or observable detail, is limited to about one wavelength. Since a potential of only 54 V can produce electrons with sub-nanometer wavelengths, it is easy to get electrons with much smaller wavelengths than those of visible light (hundreds of nanometers). Electron microscopes can, thus, be constructed to detect much smaller details than optical microscopes. (See [\[link\]](#).)

There are basically two types of electron microscopes. The transmission electron microscope (TEM) accelerates electrons that are emitted from a hot filament (the cathode). The beam is broadened and then passes through the sample. A magnetic lens focuses the beam image onto a fluorescent screen, a photographic plate, or (most probably) a CCD (light sensitive camera), from which it is transferred to a computer. The TEM is similar to the optical microscope, but it requires a thin sample examined in a vacuum. However it can resolve details as small as 0.1 nm (10^{-10} m), providing magnifications

of 100 million times the size of the original object. The TEM has allowed us to see individual atoms and structure of cell nuclei.

The scanning electron microscope (SEM) provides images by using secondary electrons produced by the primary beam interacting with the surface of the sample (see [\[link\]](#)). The SEM also uses magnetic lenses to focus the beam onto the sample. However, it moves the beam around electrically to “scan” the sample in the x and y directions. A CCD detector is used to process the data for each electron position, producing images like the one at the beginning of this chapter. The SEM has the advantage of not requiring a thin sample and of providing a 3-D view. However, its resolution is about ten times less than a TEM.



Schematic of a scanning electron microscope (SEM) (a) used to observe small details, such as those seen in this image of a tooth of a *Himipristis*, a type of shark (b). (credit: Dallas Krentzel, Flickr)

Electrons were the first particles with mass to be directly confirmed to have the wavelength proposed by de Broglie. Subsequently, protons, helium nuclei, neutrons, and many others have been observed to exhibit

interference when they interact with objects having sizes similar to their de Broglie wavelength. The de Broglie wavelength for massless particles was well established in the 1920s for photons, and it has since been observed that all massless particles have a de Broglie wavelength $\lambda = h/p$. The wave nature of all particles is a universal characteristic of nature. We shall see in following sections that implications of the de Broglie wavelength include the quantization of energy in atoms and molecules, and an alteration of our basic view of nature on the microscopic scale. The next section, for example, shows that there are limits to the precision with which we may make predictions, regardless of how hard we try. There are even limits to the precision with which we may measure an object's location or energy.

Note:

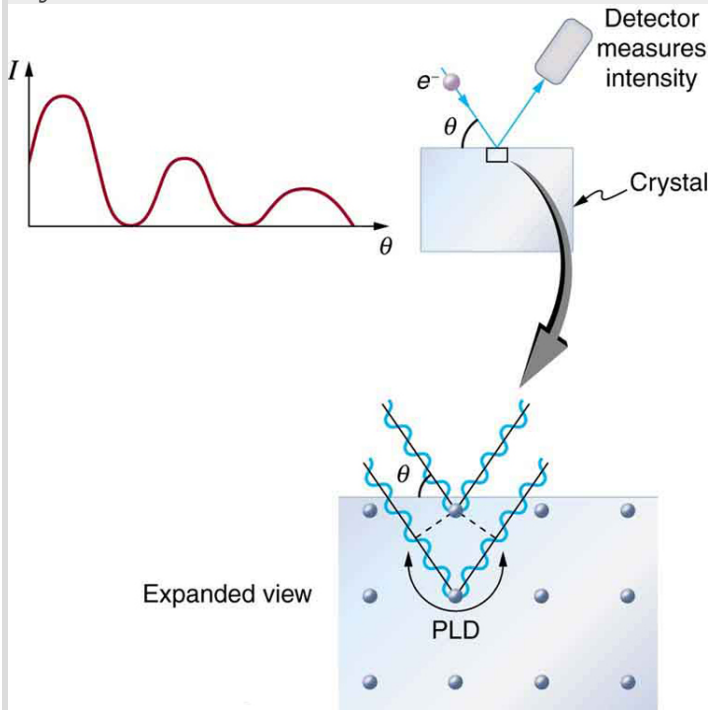
Making Connections: A Submicroscopic Diffraction Grating

The wave nature of matter allows it to exhibit all the characteristics of other, more familiar, waves. Diffraction gratings, for example, produce diffraction patterns for light that depend on grating spacing and the wavelength of the light. This effect, as with most wave phenomena, is most pronounced when the wave interacts with objects having a size similar to its wavelength. For gratings, this is the spacing between multiple slits.)

When electrons interact with a system having a spacing similar to the electron wavelength, they show the same types of interference patterns as light does for diffraction gratings, as shown at top left in [\[link\]](#).

Atoms are spaced at regular intervals in a crystal as parallel planes, as shown in the bottom part of [\[link\]](#). The spacings between these planes act like the openings in a diffraction grating. At certain incident angles, the paths of electrons scattering from successive planes differ by one wavelength and, thus, interfere constructively. At other angles, the path length differences are not an integral wavelength, and there is partial to total destructive interference. This type of scattering from a large crystal with well-defined lattice planes can produce dramatic interference patterns. It is called *Bragg reflection*, for the father-and-son team who first explored and analyzed it in some detail. The expanded view also shows the path-length differences and indicates how these depend on incident angle θ in a

manner similar to the diffraction patterns for x rays reflecting from a crystal.



The diffraction pattern at top left is produced by scattering electrons from a crystal and is graphed as a function of incident angle relative to the regular array of atoms in a crystal, as shown at bottom. Electrons scattering from the second layer of atoms travel farther than those scattered from the top layer. If the path length difference (PLD) is an integral wavelength, there is constructive interference.

Let us take the spacing between parallel planes of atoms in the crystal to be d . As mentioned, if the path length difference (PLD) for the electrons is a whole number of wavelengths, there will be constructive interference—that is, $\text{PLD} = n\lambda (n = 1, 2, 3, \dots)$. Because $AB = BC = d \sin \theta$, we have constructive interference when $n\lambda = 2d \sin \theta$. This relationship is

called the *Bragg equation* and applies not only to electrons but also to x rays.

The wavelength of matter is a submicroscopic characteristic that explains a macroscopic phenomenon such as Bragg reflection. Similarly, the wavelength of light is a submicroscopic characteristic that explains the macroscopic phenomenon of diffraction patterns.

Section Summary

- Particles of matter also have a wavelength, called the de Broglie wavelength, given by $\lambda = \frac{h}{p}$, where p is momentum.
- Matter is found to have the same *interference characteristics* as any other wave.

Conceptual Questions

Exercise:

Problem:

How does the interference of water waves differ from the interference of electrons? How are they analogous?

Exercise:

Problem: Describe one type of evidence for the wave nature of matter.

Exercise:

Problem:

Describe one type of evidence for the particle nature of EM radiation.

Problems & Exercises

Exercise:

Problem:

At what velocity will an electron have a wavelength of 1.00 m?

Solution:

$$7.28 \times 10^{-4} \text{ m}$$

Exercise:**Problem:**

What is the wavelength of an electron moving at 3.00% of the speed of light?

Exercise:**Problem:**

At what velocity does a proton have a 6.00-fm wavelength (about the size of a nucleus)? Assume the proton is nonrelativistic. (1 femtometer = 10^{-15} m.)

Solution:

$$6.62 \times 10^7 \text{ m/s}$$

Exercise:**Problem:**

What is the velocity of a 0.400-kg billiard ball if its wavelength is 7.50 cm (large enough for it to interfere with other billiard balls)?

Exercise:**Problem:**

Find the wavelength of a proton moving at 1.00% of the speed of light.

Solution:

$$1.32 \times 10^{-13} \text{ m}$$

Exercise:**Problem:**

Experiments are performed with ultracold neutrons having velocities as small as 1.00 m/s. (a) What is the wavelength of such a neutron? (b) What is its kinetic energy in eV?

Exercise:**Problem:**

(a) Find the velocity of a neutron that has a 6.00-fm wavelength (about the size of a nucleus). Assume the neutron is nonrelativistic. (b) What is the neutron's kinetic energy in MeV?

Solution:

(a) $6.62 \times 10^7 \text{ m/s}$

(b) 22.9 MeV

Exercise:**Problem:**

What is the wavelength of an electron accelerated through a 30.0-kV potential, as in a TV tube?

Exercise:**Problem:**

What is the kinetic energy of an electron in a TEM having a 0.0100-nm wavelength?

Solution:

Equation: 15.1 keV

Exercise:

Problem:

(a) Calculate the velocity of an electron that has a wavelength of $1.00\text{ }\mu\text{m}$. (b) Through what voltage must the electron be accelerated to have this velocity?

Exercise:**Problem:**

The velocity of a proton emerging from a Van de Graaff accelerator is 25.0% of the speed of light. (a) What is the proton's wavelength? (b) What is its kinetic energy, assuming it is nonrelativistic? (c) What was the equivalent voltage through which it was accelerated?

Solution:

(a) 5.29 fm

(b) $4.70 \times 10^{-12}\text{ J}$

(c) 29.4 MV

Exercise:**Problem:**

The kinetic energy of an electron accelerated in an x-ray tube is 100 keV . Assuming it is nonrelativistic, what is its wavelength?

Exercise:**Problem: Unreasonable Results**

(a) Assuming it is nonrelativistic, calculate the velocity of an electron with a 0.100-fm wavelength (small enough to detect details of a nucleus). (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $7.28 \times 10^{12} \text{ m/s}$

(b) This is thousands of times the speed of light (an impossibility).

(c) The assumption that the electron is non-relativistic is unreasonable at this wavelength.

Glossary

de Broglie wavelength

the wavelength possessed by a particle of matter, calculated by

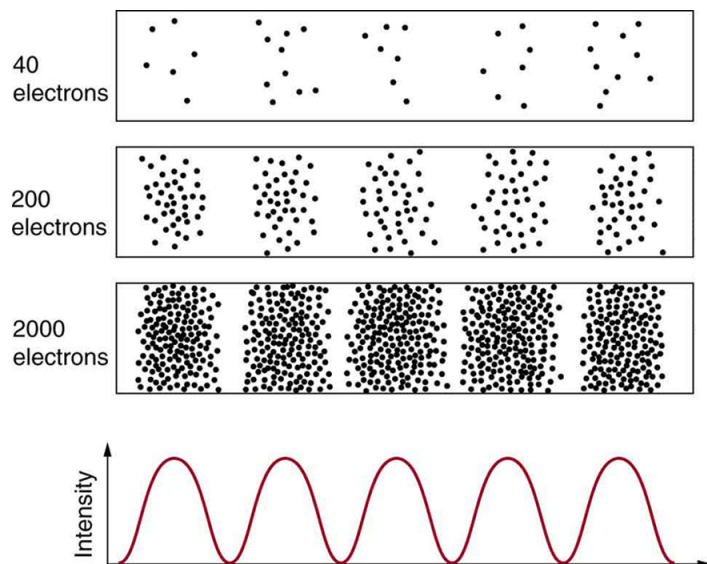
$$\lambda = h/p$$

Probability: The Heisenberg Uncertainty Principle

- Use both versions of Heisenberg's uncertainty principle in calculations.
- Explain the implications of Heisenberg's uncertainty principle for measurements.

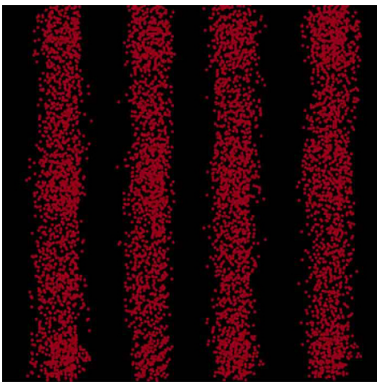
Probability Distribution

Matter and photons are waves, implying they are spread out over some distance. What is the position of a particle, such as an electron? Is it at the center of the wave? The answer lies in how you measure the position of an electron. Experiments show that you will find the electron at some definite location, unlike a wave. But if you set up exactly the same situation and measure it again, you will find the electron in a different location, often far outside any experimental uncertainty in your measurement. Repeated measurements will display a statistical distribution of locations that appears wavelike. (See [\[link\]](#).)

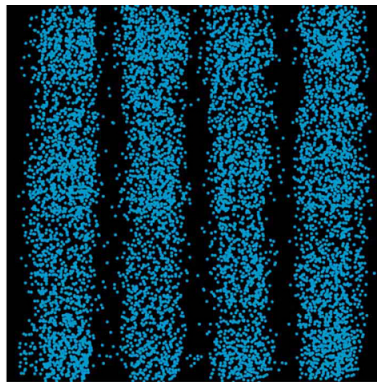


The building up of the diffraction pattern of electrons scattered from a crystal surface. Each electron arrives

at a definite location, which cannot be precisely predicted. The overall distribution shown at the bottom can be predicted as the diffraction of waves having the de Broglie wavelength of the electrons.



(a) Electrons



(b) Protons

Double-slit interference for electrons (a) and protons (b) is identical for equal wavelengths and equal slit separations. Both patterns are probability distributions in the sense that they are built up by individual particles traversing the apparatus, the paths of which are not individually predictable.

After de Broglie proposed the wave nature of matter, many physicists, including Schrödinger and Heisenberg, explored the consequences. The idea quickly emerged that, *because of its wave character, a particle's trajectory and destination cannot be precisely predicted for each particle individually*. However, each particle goes to a definite place (as illustrated in [\[link\]](#)). After compiling enough data, you get a distribution related to the

particle's wavelength and diffraction pattern. There is a certain *probability* of finding the particle at a given location, and the overall pattern is called a **probability distribution**. Those who developed quantum mechanics devised equations that predicted the probability distribution in various circumstances.

It is somewhat disquieting to think that you cannot predict exactly where an individual particle will go, or even follow it to its destination. Let us explore what happens if we try to follow a particle. Consider the double-slit patterns obtained for electrons and photons in [\[link\]](#). First, we note that these patterns are identical, following $d \sin \theta = m\lambda$, the equation for double-slit constructive interference developed in [Photon Energies and the Electromagnetic Spectrum](#), where d is the slit separation and λ is the electron or photon wavelength.

Both patterns build up statistically as individual particles fall on the detector. This can be observed for photons or electrons—for now, let us concentrate on electrons. You might imagine that the electrons are interfering with one another as any waves do. To test this, you can lower the intensity until there is never more than one electron between the slits and the screen. The same interference pattern builds up! This implies that a particle's probability distribution spans both slits, and the particles actually interfere with themselves. Does this also mean that the electron goes through both slits? An electron is a basic unit of matter that is not divisible. But it is a fair question, and so we should look to see if the electron traverses one slit or the other, or both. One possibility is to have coils around the slits that detect charges moving through them. What is observed is that an electron always goes through one slit or the other; it does not split to go through both. But there is a catch. If you determine that the electron went through one of the slits, you no longer get a double slit pattern—instead, you get single slit interference. There is no escape by using another method of determining which slit the electron went through. Knowing the particle went through one slit forces a single-slit pattern. If you do not observe which slit the electron goes through, you obtain a double-slit pattern.

Heisenberg Uncertainty

How does knowing which slit the electron passed through change the pattern? The answer is fundamentally important—*measurement affects the system being observed*. Information can be lost, and in some cases it is impossible to measure two physical quantities simultaneously to exact precision. For example, you can measure the position of a moving electron by scattering light or other electrons from it. Those probes have momentum themselves, and by scattering from the electron, they change its momentum *in a manner that loses information*. There is a limit to absolute knowledge, even in principle.



Werner Heisenberg was one of the best of those physicists who developed early quantum mechanics. Not only did his work enable a description of nature on the very small scale, it also changed our

view of the
availability of
knowledge.
Although he is
universally
recognized for his
brilliance and the
importance of his
work (he received
the Nobel Prize in
1932, for example),
Heisenberg
remained in
Germany during
World War II and
headed the German
effort to build a
nuclear bomb,
permanently
alienating himself
from most of the
scientific
community. (credit:
Author Unknown,
via Wikimedia
Commons)

It was Werner Heisenberg who first stated this limit to knowledge in 1929 as a result of his work on quantum mechanics and the wave characteristics of all particles. (See [\[link\]](#)). Specifically, consider simultaneously measuring the position and momentum of an electron (it could be any particle). There is an **uncertainty in position** Δx that is approximately equal to the wavelength of the particle. That is,
Equation:

$$\Delta x \approx \lambda.$$

As discussed above, a wave is not located at one point in space. If the electron's position is measured repeatedly, a spread in locations will be observed, implying an uncertainty in position Δx . To detect the position of the particle, we must interact with it, such as having it collide with a detector. In the collision, the particle will lose momentum. This change in momentum could be anywhere from close to zero to the total momentum of the particle, $p = h/\lambda$. It is not possible to tell how much momentum will be transferred to a detector, and so there is an **uncertainty in momentum** Δp , too. In fact, the uncertainty in momentum may be as large as the momentum itself, which in equation form means that

Equation:

$$\Delta p \approx \frac{h}{\lambda}.$$

The uncertainty in position can be reduced by using a shorter-wavelength electron, since $\Delta x \approx \lambda$. But shortening the wavelength increases the uncertainty in momentum, since $\Delta p \approx h/\lambda$. Conversely, the uncertainty in momentum can be reduced by using a longer-wavelength electron, but this increases the uncertainty in position. Mathematically, you can express this trade-off by multiplying the uncertainties. The wavelength cancels, leaving

Equation:

$$\Delta x \Delta p \approx h.$$

So if one uncertainty is reduced, the other must increase so that their product is $\approx h$.

With the use of advanced mathematics, Heisenberg showed that the best that can be done in a *simultaneous measurement of position and momentum* is

Equation:

$$\Delta x \Delta p \geq \frac{h}{4\pi}.$$

This is known as the **Heisenberg uncertainty principle**. It is impossible to measure position x and momentum p simultaneously with uncertainties Δx and Δp that multiply to be less than $h/4\pi$. Neither uncertainty can be zero. Neither uncertainty can become small without the other becoming large. A small wavelength allows accurate position measurement, but it increases the momentum of the probe to the point that it further disturbs the momentum of a system being measured. For example, if an electron is scattered from an atom and has a wavelength small enough to detect the position of electrons in the atom, its momentum can knock the electrons from their orbits in a manner that loses information about their original motion. It is therefore impossible to follow an electron in its orbit around an atom. If you measure the electron's position, you will find it in a definite location, but the atom will be disrupted. Repeated measurements on identical atoms will produce interesting probability distributions for electrons around the atom, but they will not produce motion information. The probability distributions are referred to as electron clouds or orbitals. The shapes of these orbitals are often shown in general chemistry texts and are discussed in [The Wave Nature of Matter Causes Quantization](#).

Example:

Heisenberg Uncertainty Principle in Position and Momentum for an Atom

(a) If the position of an electron in an atom is measured to an accuracy of 0.0100 nm, what is the electron's uncertainty in velocity? (b) If the electron has this velocity, what is its kinetic energy in eV?

Strategy

The uncertainty in position is the accuracy of the measurement, or $\Delta x = 0.0100$ nm. Thus the smallest uncertainty in momentum Δp can be calculated using $\Delta x \Delta p \geq h/4\pi$. Once the uncertainty in momentum Δp is found, the uncertainty in velocity can be found from $\Delta p = m\Delta v$.

Solution for (a)

Using the equals sign in the uncertainty principle to express the minimum uncertainty, we have

Equation:

$$\Delta x \Delta p = \frac{h}{4\pi}.$$

Solving for Δp and substituting known values gives

Equation:

$$\Delta p = \frac{h}{4\pi\Delta x} = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{4\pi(1.00 \times 10^{-11} \text{ m})} = 5.28 \times 10^{-24} \text{ kg} \cdot \text{m/s}.$$

Thus,

Equation:

$$\Delta p = 5.28 \times 10^{-24} \text{ kg} \cdot \text{m/s} = m\Delta v.$$

Solving for Δv and substituting the mass of an electron gives

Equation:

$$\Delta v = \frac{\Delta p}{m} = \frac{5.28 \times 10^{-24} \text{ kg} \cdot \text{m/s}}{9.11 \times 10^{-31} \text{ kg}} = 5.79 \times 10^6 \text{ m/s}.$$

Solution for (b)

Although large, this velocity is not highly relativistic, and so the electron's kinetic energy is

Equation:

$$\begin{aligned} \text{KE}_e &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}(9.11 \times 10^{-31} \text{ kg})(5.79 \times 10^6 \text{ m/s})^2 \\ &= (1.53 \times 10^{-17} \text{ J})\left(\frac{1 \text{ eV}}{1.60 \times 10^{-19} \text{ J}}\right) = 95.5 \text{ eV}. \end{aligned}$$

Discussion

Since atoms are roughly 0.1 nm in size, knowing the position of an electron to 0.0100 nm localizes it reasonably well inside the atom. This

would be like being able to see details one-tenth the size of the atom. But the consequent uncertainty in velocity is large. You certainly could not follow it very well if its velocity is so uncertain. To get a further idea of how large the uncertainty in velocity is, we assumed the velocity of the electron was equal to its uncertainty and found this gave a kinetic energy of 95.5 eV. This is significantly greater than the typical energy difference between levels in atoms (see [\[link\]](#)), so that it is impossible to get a meaningful energy for the electron if we know its position even moderately well.

Why don't we notice Heisenberg's uncertainty principle in everyday life? The answer is that Planck's constant is very small. Thus the lower limit in the uncertainty of measuring the position and momentum of large objects is negligible. We can detect sunlight reflected from Jupiter and follow the planet in its orbit around the Sun. The reflected sunlight alters the momentum of Jupiter and creates an uncertainty in its momentum, but this is totally negligible compared with Jupiter's huge momentum. The correspondence principle tells us that the predictions of quantum mechanics become indistinguishable from classical physics for large objects, which is the case here.

Heisenberg Uncertainty for Energy and Time

There is another form of **Heisenberg's uncertainty principle** for *simultaneous measurements of energy and time*. In equation form,

Equation:

$$\Delta E \Delta t \geq \frac{h}{4\pi},$$

where ΔE is the **uncertainty in energy** and Δt is the **uncertainty in time**. This means that within a time interval Δt , it is not possible to measure energy precisely—there will be an uncertainty ΔE in the measurement. In order to measure energy more precisely (to make ΔE smaller), we must

increase Δt . This time interval may be the amount of time we take to make the measurement, or it could be the amount of time a particular state exists, as in the next [\[link\]](#).

Example:**Heisenberg Uncertainty Principle for Energy and Time for an Atom**

An atom in an excited state temporarily stores energy. If the lifetime of this excited state is measured to be 1.0×10^{-10} s, what is the minimum uncertainty in the energy of the state in eV?

Strategy

The minimum uncertainty in energy ΔE is found by using the equals sign in $\Delta E \Delta t \geq h/4\pi$ and corresponds to a reasonable choice for the uncertainty in time. The largest the uncertainty in time can be is the full lifetime of the excited state, or $\Delta t = 1.0 \times 10^{-10}$ s.

Solution

Solving the uncertainty principle for ΔE and substituting known values gives

Equation:

$$\Delta E = \frac{h}{4\pi\Delta t} = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{4\pi(1.0 \times 10^{-10} \text{ s})} = 5.3 \times 10^{-25} \text{ J}.$$

Now converting to eV yields

Equation:

$$\Delta E = (5.3 \times 10^{-25} \text{ J}) \left(\frac{1 \text{ eV}}{1.6 \times 10^{-19} \text{ J}} \right) = 3.3 \times 10^{-6} \text{ eV}.$$

Discussion

The lifetime of 10^{-10} s is typical of excited states in atoms—on human time scales, they quickly emit their stored energy. An uncertainty in energy of only a few millionths of an eV results. This uncertainty is small compared with typical excitation energies in atoms, which are on the order of 1 eV. So here the uncertainty principle limits the accuracy with which

we can measure the lifetime and energy of such states, but not very significantly.

The uncertainty principle for energy and time can be of great significance if the lifetime of a system is very short. Then Δt is very small, and ΔE is consequently very large. Some nuclei and exotic particles have extremely short lifetimes (as small as 10^{-25} s), causing uncertainties in energy as great as many GeV (10^9 eV). Stored energy appears as increased rest mass, and so this means that there is significant uncertainty in the rest mass of short-lived particles. When measured repeatedly, a spread of masses or decay energies are obtained. The spread is ΔE . You might ask whether this uncertainty in energy could be avoided by not measuring the lifetime. The answer is no. Nature knows the lifetime, and so its brevity affects the energy of the particle. This is so well established experimentally that the uncertainty in decay energy is used to calculate the lifetime of short-lived states. Some nuclei and particles are so short-lived that it is difficult to measure their lifetime. But if their decay energy can be measured, its spread is ΔE , and this is used in the uncertainty principle ($\Delta E \Delta t \geq h/4\pi$) to calculate the lifetime Δt .

There is another consequence of the uncertainty principle for energy and time. If energy is uncertain by ΔE , then conservation of energy can be violated by ΔE for a time Δt . Neither the physicist nor nature can tell that conservation of energy has been violated, if the violation is temporary and smaller than the uncertainty in energy. While this sounds innocuous enough, we shall see in later chapters that it allows the temporary creation of matter from nothing and has implications for how nature transmits forces over very small distances.

Finally, note that in the discussion of particles and waves, we have stated that individual measurements produce precise or particle-like results. A definite position is determined each time we observe an electron, for example. But repeated measurements produce a spread in values consistent with wave characteristics. The great theoretical physicist Richard Feynman (1918–1988) commented, “What there are, are particles.” When you

observe enough of them, they distribute themselves as you would expect for a wave phenomenon. However, what there are as they travel we cannot tell because, when we do try to measure, we affect the traveling.

Section Summary

- Matter is found to have the same interference characteristics as any other wave.
- There is now a probability distribution for the location of a particle rather than a definite position.
- Another consequence of the wave character of all particles is the Heisenberg uncertainty principle, which limits the precision with which certain physical quantities can be known simultaneously. For position and momentum, the uncertainty principle is $\Delta x \Delta p \geq \frac{h}{4\pi}$, where Δx is the uncertainty in position and Δp is the uncertainty in momentum.
- For energy and time, the uncertainty principle is $\Delta E \Delta t \geq \frac{h}{4\pi}$ where ΔE is the uncertainty in energy and Δt is the uncertainty in time.
- These small limits are fundamentally important on the quantum-mechanical scale.

Conceptual Questions

Exercise:

Problem:

What is the Heisenberg uncertainty principle? Does it place limits on what can be known?

Problems & Exercises

Exercise:

Problem:

(a) If the position of an electron in a membrane is measured to an accuracy of $1.00\ \mu\text{m}$, what is the electron's minimum uncertainty in velocity? (b) If the electron has this velocity, what is its kinetic energy in eV? (c) What are the implications of this energy, comparing it to typical molecular binding energies?

Solution:

(a) $57.9\ \text{m/s}$

(b) $9.55 \times 10^{-9}\ \text{eV}$

(c) From [\[link\]](#), we see that typical molecular binding energies range from about 1 eV to 10 eV, therefore the result in part (b) is approximately 9 orders of magnitude smaller than typical molecular binding energies.

Exercise:**Problem:**

(a) If the position of a chlorine ion in a membrane is measured to an accuracy of $1.00\ \mu\text{m}$, what is its minimum uncertainty in velocity, given its mass is $5.86 \times 10^{-26}\ \text{kg}$? (b) If the ion has this velocity, what is its kinetic energy in eV, and how does this compare with typical molecular binding energies?

Exercise:**Problem:**

Suppose the velocity of an electron in an atom is known to an accuracy of $2.0 \times 10^3\ \text{m/s}$ (reasonably accurate compared with orbital velocities). What is the electron's minimum uncertainty in position, and how does this compare with the approximate 0.1-nm size of the atom?

Solution:

29 nm,

290 times greater

Exercise:**Problem:**

The velocity of a proton in an accelerator is known to an accuracy of 0.250% of the speed of light. (This could be small compared with its velocity.) What is the smallest possible uncertainty in its position?

Exercise:**Problem:**

A relatively long-lived excited state of an atom has a lifetime of 3.00 ms. What is the minimum uncertainty in its energy?

Solution:

$$1.10 \times 10^{-13} \text{ eV}$$

Exercise:**Problem:**

(a) The lifetime of a highly unstable nucleus is 10^{-20} s. What is the smallest uncertainty in its decay energy? (b) Compare this with the rest energy of an electron.

Exercise:**Problem:**

The decay energy of a short-lived particle has an uncertainty of 1.0 MeV due to its short lifetime. What is the smallest lifetime it can have?

Solution:

$$3.3 \times 10^{-22} \text{ s}$$

Exercise:**Problem:**

The decay energy of a short-lived nuclear excited state has an uncertainty of 2.0 eV due to its short lifetime. What is the smallest lifetime it can have?

Exercise:**Problem:**

What is the approximate uncertainty in the mass of a muon, as determined from its decay lifetime?

Solution:

$$2.66 \times 10^{-46} \text{ kg}$$

Exercise:**Problem:**

Derive the approximate form of Heisenberg's uncertainty principle for energy and time, $\Delta E \Delta t \approx h$, using the following arguments: Since the position of a particle is uncertain by $\Delta x \approx \lambda$, where λ is the wavelength of the photon used to examine it, there is an uncertainty in the time the photon takes to traverse Δx . Furthermore, the photon has an energy related to its wavelength, and it can transfer some or all of this energy to the object being examined. Thus the uncertainty in the energy of the object is also related to λ . Find Δt and ΔE ; then multiply them to give the approximate uncertainty principle.

Glossary**Heisenberg's uncertainty principle**

a fundamental limit to the precision with which pairs of quantities (momentum and position, and energy and time) can be measured

uncertainty in energy

lack of precision or lack of knowledge of precise results in measurements of energy

uncertainty in time

lack of precision or lack of knowledge of precise results in measurements of time

uncertainty in momentum

lack of precision or lack of knowledge of precise results in measurements of momentum

uncertainty in position

lack of precision or lack of knowledge of precise results in measurements of position

probability distribution

the overall spatial distribution of probabilities to find a particle at a given location

The Particle-Wave Duality Reviewed

- Explain the concept of particle-wave duality, and its scope.

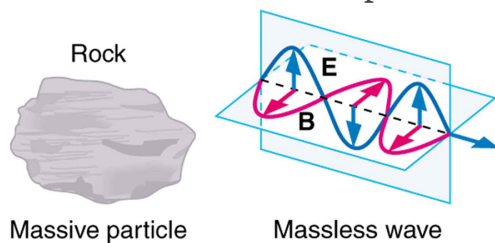
Particle-wave duality—the fact that all particles have wave properties—is one of the cornerstones of quantum mechanics. We first came across it in the treatment of photons, those particles of EM radiation that exhibit both particle and wave properties, but not at the same time. Later it was noted that particles of matter have wave properties as well. The dual properties of particles and waves are found for all particles, whether massless like photons, or having a mass like electrons. (See [\[link\]](#).)



On a quantum-mechanical scale (i.e., very small), particles with and without mass have wave properties. For example, both electrons and photons have wavelengths but also behave as particles.

There are many submicroscopic particles in nature. Most have mass and are expected to act as particles, or the smallest units of matter. All these masses have wave properties, with wavelengths given by the de Broglie relationship $\lambda = h/p$. So, too, do combinations of these particles, such as nuclei, atoms, and molecules. As a combination of masses becomes large, particularly if it is large enough to be called macroscopic, its wave nature becomes difficult to observe. This is consistent with our common experience with matter.

Some particles in nature are massless. We have only treated the photon so far, but all massless entities travel at the speed of light, have a wavelength, and exhibit particle and wave behaviors. They have momentum given by a rearrangement of the de Broglie relationship, $p = h/\lambda$. In large combinations of these massless particles (such large combinations are common only for photons or EM waves), there is mostly wave behavior upon detection, and the particle nature becomes difficult to observe. This is also consistent with experience. (See [\[link\]](#).)



On a classical scale (macroscopic), particles with mass behave as particles and not as waves. Particles without mass act as waves and not as particles.

The particle-wave duality is a universal attribute. It is another connection between matter and energy. Not only has modern physics been able to

describe nature for high speeds and small sizes, it has also discovered new connections and symmetries. There is greater unity and symmetry in nature than was known in the classical era—but they were dreamt of. A beautiful poem written by the English poet William Blake some two centuries ago contains the following four lines:

To see the World in a Grain of Sand

And a Heaven in a Wild Flower

Hold Infinity in the palm of your hand

And Eternity in an hour

Integrated Concepts

The problem set for this section involves concepts from this chapter and several others. Physics is most interesting when applied to general situations involving more than a narrow set of physical principles. For example, photons have momentum, hence the relevance of [Linear Momentum and Collisions](#). The following topics are involved in some or all of the problems in this section:

- [Dynamics: Newton's Laws of Motion](#)
- [Work, Energy, and Energy Resources](#)
- [Linear Momentum and Collisions](#)
- [Heat and Heat Transfer Methods](#)
- [Electric Potential and Electric Field](#)
- [Electric Current, Resistance, and Ohm's Law](#)
- [Wave Optics](#)
- [Special Relativity](#)

Note:

Problem-Solving Strategy

1. Identify which physical principles are involved.

2. Solve the problem using strategies outlined in the text.

[\[link\]](#) illustrates how these strategies are applied to an integrated-concept problem.

Example:

Recoil of a Dust Particle after Absorbing a Photon

The following topics are involved in this integrated concepts worked example:

Photons (quantum mechanics)

Linear Momentum

Topics

A 550-nm photon (visible light) is absorbed by a 1.00- μg particle of dust in outer space. (a) Find the momentum of such a photon. (b) What is the recoil velocity of the particle of dust, assuming it is initially at rest?

Strategy Step 1

To solve an *integrated-concept problem*, such as those following this example, we must first identify the physical principles involved and identify the chapters in which they are found. Part (a) of this example asks for the *momentum of a photon*, a topic of the present chapter. Part (b) considers *recoil following a collision*, a topic of [Linear Momentum and Collisions](#).

Strategy Step 2

The following solutions to each part of the example illustrate how specific problem-solving strategies are applied. These involve identifying knowns and unknowns, checking to see if the answer is reasonable, and so on.

Solution for (a)

The momentum of a photon is related to its wavelength by the equation:

Equation:

$$p = \frac{h}{\lambda}.$$

Entering the known value for Planck's constant h and given the wavelength λ , we obtain

Equation:

$$\begin{aligned} p &= \frac{6.63 \times 10^{-34} \text{ J}\cdot\text{s}}{550 \times 10^{-9} \text{ m}} \\ &= 1.21 \times 10^{-27} \text{ kg} \cdot \text{m/s}. \end{aligned}$$

Discussion for (a)

This momentum is small, as expected from discussions in the text and the fact that photons of visible light carry small amounts of energy and momentum compared with those carried by macroscopic objects.

Solution for (b)

Conservation of momentum in the absorption of this photon by a grain of dust can be analyzed using the equation:

Equation:

$$p_1 + p_2 = p'_1 + p'_2 (F_{\text{net}} = 0).$$

The net external force is zero, since the dust is in outer space. Let 1 represent the photon and 2 the dust particle. Before the collision, the dust is at rest (relative to some observer); after the collision, there is no photon (it is absorbed). So conservation of momentum can be written

Equation:

$$p_1 = p'_2 = mv,$$

where p_1 is the photon momentum before the collision and p'_2 is the dust momentum after the collision. The mass and recoil velocity of the dust are

m and v , respectively. Solving this for v , the requested quantity, yields

Equation:

$$v = \frac{p}{m},$$

where p is the photon momentum found in part (a). Entering known values (noting that a microgram is 10^{-9} kg) gives

Equation:

$$\begin{aligned} v &= \frac{1.21 \times 10^{-27} \text{ kg}\cdot\text{m/s}}{1.00 \times 10^{-9} \text{ kg}} \\ &= 1.21 \times 10^{-18} \text{ m/s.} \end{aligned}$$

Discussion

The recoil velocity of the particle of dust is extremely small. As we have noted, however, there are immense numbers of photons in sunlight and other macroscopic sources. In time, collisions and absorption of many photons could cause a significant recoil of the dust, as observed in comet tails.

Section Summary

- The particle-wave duality refers to the fact that all particles—those with mass and those without mass—have wave characteristics.
- This is a further connection between mass and energy.

Conceptual Questions

Exercise:

Problem:

In what ways are matter and energy related that were not known before the development of relativity and quantum mechanics?

Problems & Exercises

Exercise:

Problem: Integrated Concepts

The 54.0-eV electron in [\[link\]](#) has a 0.167-nm wavelength. If such electrons are passed through a double slit and have their first maximum at an angle of 25.0° , what is the slit separation d ?

Solution:

0.395 nm

Exercise:

Problem: Integrated Concepts

An electron microscope produces electrons with a 2.00-pm wavelength. If these are passed through a 1.00-nm single slit, at what angle will the first diffraction minimum be found?

Exercise:

Problem: Integrated Concepts

A certain heat lamp emits 200 W of mostly IR radiation averaging 1500 nm in wavelength. (a) What is the average photon energy in joules? (b) How many of these photons are required to increase the temperature of a person's shoulder by 2.0°C , assuming the affected mass is 4.0 kg with a specific heat of $0.83 \text{ kcal/kg}\cdot^\circ\text{C}$. Also assume no other significant heat transfer. (c) How long does this take?

Solution:

(a) $1.3 \times 10^{-19} \text{ J}$

(b) 2.1×10^{23}

(c) $1.4 \times 10^2 \text{ s}$

Exercise:

Problem: Integrated Concepts

On its high power setting, a microwave oven produces 900 W of 2560 MHz microwaves. (a) How many photons per second is this? (b) How many photons are required to increase the temperature of a 0.500-kg mass of pasta by 45.0°C , assuming a specific heat of $0.900 \text{ kcal/kg} \cdot ^\circ\text{C}$? Neglect all other heat transfer. (c) How long must the microwave operator wait for their pasta to be ready?

Exercise:

Problem: Integrated Concepts

(a) Calculate the amount of microwave energy in joules needed to raise the temperature of 1.00 kg of soup from 20.0°C to 100°C . (b) What is the total momentum of all the microwave photons it takes to do this? (c) Calculate the velocity of a 1.00-kg mass with the same momentum. (d) What is the kinetic energy of this mass?

Solution:

(a) $3.35 \times 10^5 \text{ J}$

(b) $1.12 \times 10^{-3} \text{ kg} \cdot \text{m/s}$

(c) $1.12 \times 10^{-3} \text{ m/s}$

(d) $6.23 \times 10^{-7} \text{ J}$

Exercise:

Problem: Integrated Concepts

- (a) What is γ for an electron emerging from the Stanford Linear Accelerator with a total energy of 50.0 GeV? (b) Find its momentum. (c) What is the electron's wavelength?

Exercise:

Problem: Integrated Concepts

- (a) What is γ for a proton having an energy of 1.00 TeV, produced by the Fermilab accelerator? (b) Find its momentum. (c) What is the proton's wavelength?

Solution:

- (a) 1.06×10^3
(b) $5.33 \times 10^{-16} \text{ kg} \cdot \text{m/s}$
(c) $1.24 \times 10^{-18} \text{ m}$

Exercise:

Problem: Integrated Concepts

An electron microscope passes 1.00-pm-wavelength electrons through a circular aperture 2.00 μm in diameter. What is the angle between two just-resolvable point sources for this microscope?

Exercise:

Problem: Integrated Concepts

- (a) Calculate the velocity of electrons that form the same pattern as 450-nm light when passed through a double slit. (b) Calculate the kinetic energy of each and compare them. (c) Would either be easier to generate than the other? Explain.

Solution:

(a) $1.62 \times 10^3 \text{ m/s}$

(b) $4.42 \times 10^{-19} \text{ J}$ for photon, $1.19 \times 10^{-24} \text{ J}$ for electron, photon energy is 3.71×10^5 times greater

(c) The light is easier to make because 450-nm light is blue light and therefore easy to make. Creating electrons with $7.43 \text{ } \mu\text{eV}$ of energy would not be difficult, but would require a vacuum.

Exercise:

Problem: Integrated Concepts

(a) What is the separation between double slits that produces a second-order minimum at 45.0° for 650-nm light? (b) What slit separation is needed to produce the same pattern for 1.00-keV protons.

Solution:

(a) $2.30 \times 10^{-6} \text{ m}$

(b) $3.20 \times 10^{-12} \text{ m}$

Exercise:

Problem: Integrated Concepts

A laser with a power output of 2.00 mW at a wavelength of 400 nm is projected onto calcium metal. (a) How many electrons per second are ejected? (b) What power is carried away by the electrons, given that the binding energy is 2.71 eV? (c) Calculate the current of ejected electrons. (d) If the photoelectric material is electrically insulated and acts like a 2.00-pF capacitor, how long will current flow before the capacitor voltage stops it?

Exercise:

Problem: Integrated Concepts

One problem with x rays is that they are not sensed. Calculate the temperature increase of a researcher exposed in a few seconds to a nearly fatal accidental dose of x rays under the following conditions. The energy of the x-ray photons is 200 keV, and 4.00×10^{13} of them are absorbed per kilogram of tissue, the specific heat of which is $0.830 \text{ kcal/kg} \cdot ^\circ\text{C}$. (Note that medical diagnostic x-ray machines *cannot* produce an intensity this great.)

Solution:

$$3.69 \times 10^{-4} \text{ }^\circ\text{C}$$

Exercise:

Problem: Integrated Concepts

A 1.00-fm photon has a wavelength short enough to detect some information about nuclei. (a) What is the photon momentum? (b) What is its energy in joules and MeV? (c) What is the (relativistic) velocity of an electron with the same momentum? (d) Calculate the electron's kinetic energy.

Exercise:

Problem: Integrated Concepts

The momentum of light is exactly reversed when reflected straight back from a mirror, assuming negligible recoil of the mirror. Thus the change in momentum is twice the photon momentum. Suppose light of intensity 1.00 kW/m^2 reflects from a mirror of area 2.00 m^2 . (a) Calculate the energy reflected in 1.00 s. (b) What is the momentum imparted to the mirror? (c) Using the most general form of Newton's second law, what is the force on the mirror? (d) Does the assumption of no mirror recoil seem reasonable?

Solution:

(a) 2.00 kJ

(b) $1.33 \times 10^{-5} \text{ kg} \cdot \text{m/s}$

(c) $1.33 \times 10^{-5} \text{ N}$

(d) yes

Exercise:

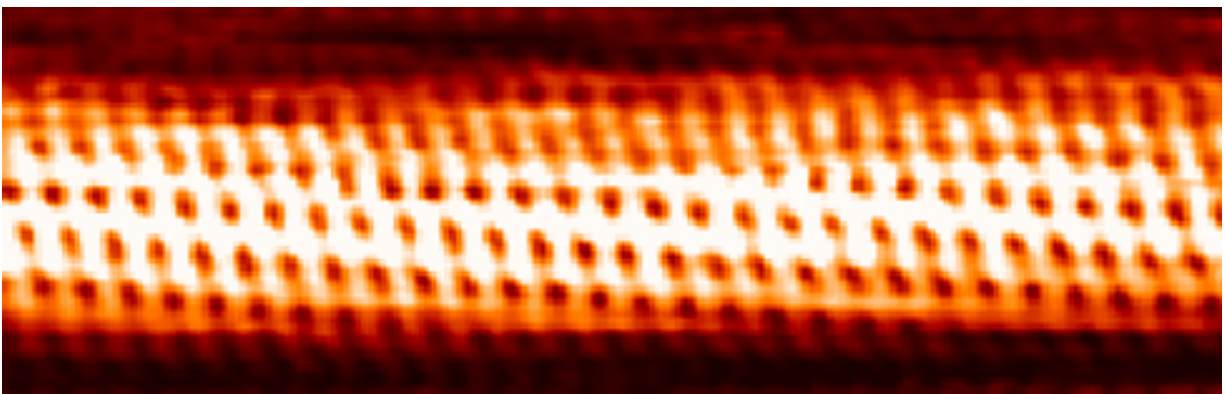
Problem: Integrated Concepts

Sunlight above the Earth's atmosphere has an intensity of 1.30 kW/m^2 . If this is reflected straight back from a mirror that has only a small recoil, the light's momentum is exactly reversed, giving the mirror twice the incident momentum. (a) Calculate the force per square meter of mirror. (b) Very low mass mirrors can be constructed in the near weightlessness of space, and attached to a spaceship to sail it. Once done, the average mass per square meter of the spaceship is 0.100 kg . Find the acceleration of the spaceship if all other forces are balanced. (c) How fast is it moving 24 hours later?

Introduction to Atomic Physics

class="introduction"

Individual
carbon
atoms are
visible in
this image
of a carbon
nanotube
made by a
scanning
tunneling
electron
microscope
. (credit:
Taner
Yildirim,
National
Institute of
Standards
and
Technology
, via
Wikimedia
Commons)



From childhood on, we learn that atoms are a substructure of all things around us, from the air we breathe to the autumn leaves that blanket a forest trail. Invisible to the eye, the existence and properties of atoms are used to explain many phenomena—a theme found throughout this text. In this chapter, we discuss the discovery of atoms and their own substructures; we then apply quantum mechanics to the description of atoms, and their properties and interactions. Along the way, we will find, much like the scientists who made the original discoveries, that new concepts emerge with applications far beyond the boundaries of atomic physics.

Discovery of the Atom

- Describe the basic structure of the atom, the substructure of all matter.

How do we know that atoms are really there if we cannot see them with our eyes? A brief account of the progression from the proposal of atoms by the Greeks to the first direct evidence of their existence follows.

People have long speculated about the structure of matter and the existence of atoms. The earliest significant ideas to survive are due to the ancient Greeks in the fifth century BCE, especially those of the philosophers Leucippus and Democritus. (There is some evidence that philosophers in both India and China made similar speculations, at about the same time.) They considered the question of whether a substance can be divided without limit into ever smaller pieces. There are only a few possible answers to this question. One is that infinitesimally small subdivision is possible. Another is what Democritus in particular believed—that there is a smallest unit that cannot be further subdivided. Democritus called this the **atom**. We now know that atoms themselves can be subdivided, but their identity is destroyed in the process, so the Greeks were correct in a respect. The Greeks also felt that atoms were in constant motion, another correct notion.

The Greeks and others speculated about the properties of atoms, proposing that only a few types existed and that all matter was formed as various combinations of these types. The famous proposal that the basic elements were earth, air, fire, and water was brilliant, but incorrect. The Greeks had identified the most common examples of the four states of matter (solid, gas, plasma, and liquid), rather than the basic elements. More than 2000 years passed before observations could be made with equipment capable of revealing the true nature of atoms.

Over the centuries, discoveries were made regarding the properties of substances and their chemical reactions. Certain systematic features were recognized, but similarities between common and rare elements resulted in efforts to transmute them (lead into gold, in particular) for financial gain. Secrecy was endemic. Alchemists discovered and rediscovered many facts but did not make them broadly available. As the Middle Ages ended, alchemy gradually faded, and the science of chemistry arose. It was no

longer possible, nor considered desirable, to keep discoveries secret. Collective knowledge grew, and by the beginning of the 19th century, an important fact was well established—the masses of reactants in specific chemical reactions always have a particular mass ratio. This is very strong indirect evidence that there are basic units (atoms and molecules) that have these same mass ratios. The English chemist John Dalton (1766–1844) did much of this work, with significant contributions by the Italian physicist Amedeo Avogadro (1776–1856). It was Avogadro who developed the idea of a fixed number of atoms and molecules in a mole, and this special number is called Avogadro's number in his honor. The Austrian physicist Johann Josef Loschmidt was the first to measure the value of the constant in 1865 using the kinetic theory of gases.

Note:**Patterns and Systematics**

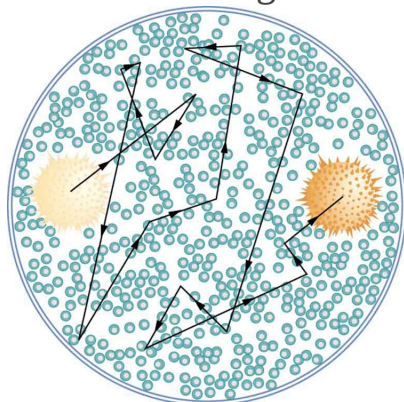
The recognition and appreciation of patterns has enabled us to make many discoveries. The periodic table of elements was proposed as an organized summary of the known elements long before all elements had been discovered, and it led to many other discoveries. We shall see in later chapters that patterns in the properties of subatomic particles led to the proposal of quarks as their underlying structure, an idea that is still bearing fruit.

Knowledge of the properties of elements and compounds grew, culminating in the mid-19th-century development of the periodic table of the elements by Dmitri Mendeleev (1834–1907), the great Russian chemist. Mendeleev proposed an ingenious array that highlighted the periodic nature of the properties of elements. Believing in the systematics of the periodic table, he also predicted the existence of then-unknown elements to complete it. Once these elements were discovered and determined to have properties predicted by Mendeleev, his periodic table became universally accepted.

Also during the 19th century, the kinetic theory of gases was developed. Kinetic theory is based on the existence of atoms and molecules in random

thermal motion and provides a microscopic explanation of the gas laws, heat transfer, and thermodynamics (see [Introduction to Temperature, Kinetic Theory, and the Gas Laws](#) and [Introduction to Laws of Thermodynamics](#)). Kinetic theory works so well that it is another strong indication of the existence of atoms. But it is still indirect evidence—individual atoms and molecules had not been observed. There were heated debates about the validity of kinetic theory until direct evidence of atoms was obtained.

The first truly direct evidence of atoms is credited to Robert Brown, a Scottish botanist. In 1827, he noticed that tiny pollen grains suspended in still water moved about in complex paths. This can be observed with a microscope for any small particles in a fluid. The motion is caused by the random thermal motions of fluid molecules colliding with particles in the fluid, and it is now called **Brownian motion**. (See [\[link\]](#).) Statistical fluctuations in the numbers of molecules striking the sides of a visible particle cause it to move first this way, then that. Although the molecules cannot be directly observed, their effects on the particle can be. By examining Brownian motion, the size of molecules can be calculated. The smaller and more numerous they are, the smaller the fluctuations in the numbers striking different sides.



The position of a
pollen grain in
water, measured
every few seconds
under a
microscope,

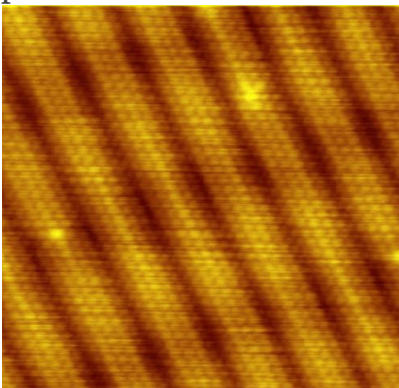
exhibits Brownian motion. Brownian motion is due to fluctuations in the number of atoms and molecules colliding with a small mass, causing it to move about in complex paths. This is nearly direct evidence for the existence of atoms, providing a satisfactory alternative explanation cannot be found.

It was Albert Einstein who, starting in his epochal year of 1905, published several papers that explained precisely how Brownian motion could be used to measure the size of atoms and molecules. (In 1905 Einstein created special relativity, proposed photons as quanta of EM radiation, and produced a theory of Brownian motion that allowed the size of atoms to be determined. All of this was done in his spare time, since he worked days as a patent examiner. Any one of these very basic works could have been the crowning achievement of an entire career—yet Einstein did even more in later years.) Their sizes were only approximately known to be 10^{-10} m, based on a comparison of latent heat of vaporization and surface tension made in about 1805 by Thomas Young of double-slit fame and the famous astronomer and mathematician Simon Laplace.

Using Einstein's ideas, the French physicist Jean-Baptiste Perrin (1870–1942) carefully observed Brownian motion; not only did he confirm Einstein's theory, he also produced accurate sizes for atoms and molecules.

Since molecular weights and densities of materials were well established, knowing atomic and molecular sizes allowed a precise value for Avogadro's number to be obtained. (If we know how big an atom is, we know how many fit into a certain volume.) Perrin also used these ideas to explain atomic and molecular agitation effects in sedimentation, and he received the 1926 Nobel Prize for his achievements. Most scientists were already convinced of the existence of atoms, but the accurate observation and analysis of Brownian motion was conclusive—it was the first truly direct evidence.

A huge array of direct and indirect evidence for the existence of atoms now exists. For example, it has become possible to accelerate ions (much as electrons are accelerated in cathode-ray tubes) and to detect them individually as well as measure their masses (see [More Applications of Magnetism](#) for a discussion of mass spectrometers). Other devices that observe individual atoms, such as the scanning tunneling electron microscope, will be discussed elsewhere. (See [\[link\]](#).) All of our understanding of the properties of matter is based on and consistent with the atom. The atom's substructures, such as electron shells and the nucleus, are both interesting and important. The nucleus in turn has a substructure, as do the particles of which it is composed. These topics, and the question of whether there is a smallest basic structure to matter, will be explored in later parts of the text.



Individual atoms
can be detected
with devices such
as the scanning
tunneling electron

microscope that
produced this
image of individual
gold atoms on a
graphite substrate.
(credit: Erwin
Rossen, Eindhoven
University of
Technology, via
Wikimedia
Commons)

Section Summary

- Atoms are the smallest unit of elements; atoms combine to form molecules, the smallest unit of compounds.
- The first direct observation of atoms was in Brownian motion.
- Analysis of Brownian motion gave accurate sizes for atoms (10^{-10} m on average) and a precise value for Avogadro's number.

Conceptual Questions

Exercise:

Problem:

Name three different types of evidence for the existence of atoms.

Exercise:

Problem:

Explain why patterns observed in the periodic table of the elements are evidence for the existence of atoms, and why Brownian motion is a more direct type of evidence for their existence.

Exercise:

Problem: If atoms exist, why can't we see them with visible light?

Problems & Exercises

Exercise:

Problem:

Using the given charge-to-mass ratios for electrons and protons, and knowing the magnitudes of their charges are equal, what is the ratio of the proton's mass to the electron's? (Note that since the charge-to-mass ratios are given to only three-digit accuracy, your answer may differ from the accepted ratio in the fourth digit.)

Solution:

$$1.84 \times 10^3$$

Exercise:

Problem:

(a) Calculate the mass of a proton using the charge-to-mass ratio given for it in this chapter and its known charge. (b) How does your result compare with the proton mass given in this chapter?

Exercise:

Problem:

If someone wanted to build a scale model of the atom with a nucleus 1.00 m in diameter, how far away would the nearest electron need to be?

Solution:

50 km

Glossary

atom

basic unit of matter, which consists of a central, positively charged nucleus surrounded by negatively charged electrons

Brownian motion

the continuous random movement of particles of matter suspended in a liquid or gas

Discovery of the Parts of the Atom: Electrons and Nuclei

- Describe how electrons were discovered.
- Explain the Millikan oil drop experiment.
- Describe Rutherford's gold foil experiment.
- Describe Rutherford's planetary model of the atom.

Just as atoms are a substructure of matter, electrons and nuclei are substructures of the atom. The experiments that were used to discover electrons and nuclei reveal some of the basic properties of atoms and can be readily understood using ideas such as electrostatic and magnetic force, already covered in previous chapters.

Note:

Charges and Electromagnetic Forces

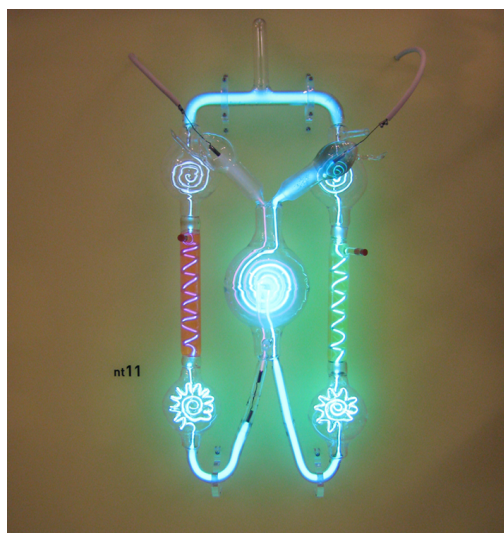
In previous discussions, we have noted that positive charge is associated with nuclei and negative charge with electrons. We have also covered many aspects of the electric and magnetic forces that affect charges. We will now explore the discovery of the electron and nucleus as substructures of the atom and examine their contributions to the properties of atoms.

The Electron

Gas discharge tubes, such as that shown in [\[link\]](#), consist of an evacuated glass tube containing two metal electrodes and a rarefied gas. When a high voltage is applied to the electrodes, the gas glows. These tubes were the precursors to today's neon lights. They were first studied seriously by Heinrich Geissler, a German inventor and glassblower, starting in the 1860s. The English scientist William Crookes, among others, continued to study what for some time were called Crookes tubes, wherein electrons are freed from atoms and molecules in the rarefied gas inside the tube and are accelerated from the cathode (negative) to the anode (positive) by the high potential. These "*cathode rays*" collide with the gas atoms and molecules and excite them, resulting in the emission of electromagnetic (EM)

radiation that makes the electrons' path visible as a ray that spreads and fades as it moves away from the cathode.

Gas discharge tubes today are most commonly called **cathode-ray tubes**, because the rays originate at the cathode. Crookes showed that the electrons carry momentum (they can make a small paddle wheel rotate). He also found that their normally straight path is bent by a magnet in the direction expected for a negative charge moving away from the cathode. These were the first direct indications of electrons and their charge.



A gas discharge tube
glows when a high
voltage is applied to it.
Electrons emitted from
the cathode are
accelerated toward the
anode; they excite atoms
and molecules in the gas,
which glow in response.
Once called Geissler
tubes and later Crookes
tubes, they are now
known as cathode-ray

tubes (CRTs) and are found in older TVs, computer screens, and x-ray machines. When a magnetic field is applied, the beam bends in the direction expected for negative charge. (credit: Paul Downey, Flickr)

The English physicist J. J. Thomson (1856–1940) improved and expanded the scope of experiments with gas discharge tubes. (See [\[link\]](#) and [\[link\]](#).) He verified the negative charge of the cathode rays with both magnetic and electric fields. Additionally, he collected the rays in a metal cup and found an excess of negative charge. Thomson was also able to measure the ratio of the charge of the electron to its mass, q_e/m_e —an important step to finding the actual values of both q_e and m_e . [\[link\]](#) shows a cathode-ray tube, which produces a narrow beam of electrons that passes through charging plates connected to a high-voltage power supply. An electric field \mathbf{E} is produced between the charging plates, and the cathode-ray tube is placed between the poles of a magnet so that the electric field \mathbf{E} is perpendicular to the magnetic field \mathbf{B} of the magnet. These fields, being perpendicular to each other, produce opposing forces on the electrons. As discussed for mass spectrometers in [More Applications of Magnetism](#), if the net force due to the fields vanishes, then the velocity of the charged particle is $v = E/B$. In this manner, Thomson determined the velocity of the electrons and then moved the beam up and down by adjusting the electric field.



J. J. Thomson (credit:
www.firstworldwar.com
, via Wikimedia
Commons)

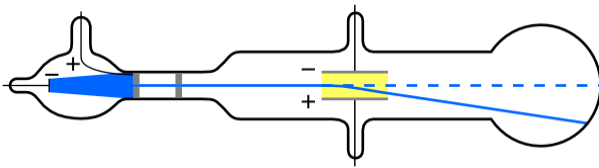
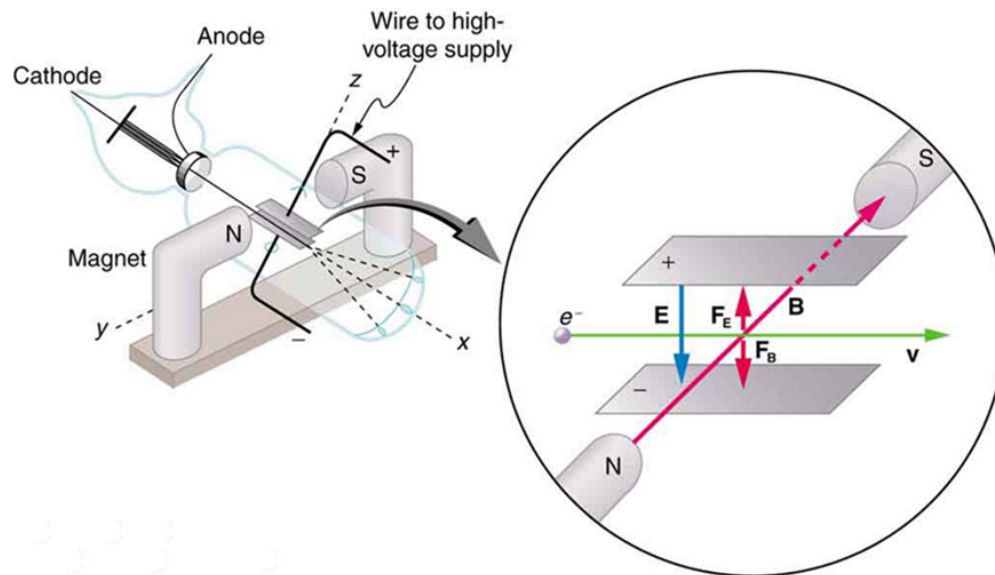


Diagram of Thomson's CRT.
(credit: Kurzon, Wikimedia
Commons)



This schematic shows the electron beam in a CRT passing through crossed electric and magnetic fields and causing phosphor to glow when striking the end of the tube.

To see how the amount of deflection is used to calculate q_e/m_e , note that the deflection is proportional to the electric force on the electron:

Equation:

$$F = q_e E.$$

But the vertical deflection is also related to the electron's mass, since the electron's acceleration is

Equation:

$$a = \frac{F}{m_e}.$$

The value of F is not known, since q_e was not yet known. Substituting the expression for electric force into the expression for acceleration yields

Equation:

$$a = \frac{F}{m_e} = \frac{q_e E}{m_e}.$$

Gathering terms, we have

Equation:

$$\frac{q_e}{m_e} = \frac{a}{E}.$$

The deflection is analyzed to get a , and E is determined from the applied voltage and distance between the plates; thus, $\frac{q_e}{m_e}$ can be determined. With the velocity known, another measurement of $\frac{q_e}{m_e}$ can be obtained by bending the beam of electrons with the magnetic field. Since $F_{\text{mag}} = q_e vB = m_e a$, we have $q_e/m_e = a/vB$. Consistent results are obtained using magnetic deflection.

What is so important about q_e/m_e , the ratio of the electron's charge to its mass? The value obtained is

Equation:

$$\frac{q_e}{m_e} = -1.76 \times 10^{11} \text{ C/kg (electron)}.$$

This is a huge number, as Thomson realized, and it implies that the electron has a very small mass. It was known from electroplating that about 10^8 C/kg is needed to plate a material, a factor of about 1000 less than the charge per kilogram of electrons. Thomson went on to do the same experiment for positively charged hydrogen ions (now known to be bare protons) and found a charge per kilogram about 1000 times smaller than that for the electron, implying that the proton is about 1000 times more massive than the electron. Today, we know more precisely that

Equation:

$$\frac{q_p}{m_p} = 9.58 \times 10^7 \text{ C/kg (proton),}$$

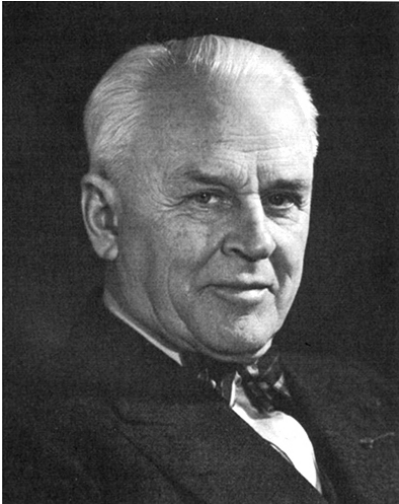
where q_p is the charge of the proton and m_p is its mass. This ratio (to four significant figures) is 1836 times less charge per kilogram than for the electron. Since the charges of electrons and protons are equal in magnitude, this implies $m_p = 1836m_e$.

Thomson performed a variety of experiments using differing gases in discharge tubes and employing other methods, such as the photoelectric effect, for freeing electrons from atoms. He always found the same properties for the electron, proving it to be an independent particle. For his work, the important pieces of which he began to publish in 1897, Thomson was awarded the 1906 Nobel Prize in Physics. In retrospect, it is difficult to appreciate how astonishing it was to find that the atom has a substructure. Thomson himself said, “It was only when I was convinced that the experiment left no escape from it that I published my belief in the existence of bodies smaller than atoms.”

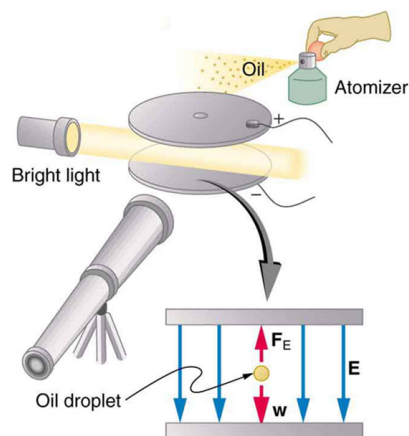
Thomson attempted to measure the charge of individual electrons, but his method could determine its charge only to the order of magnitude expected.

Since Faraday’s experiments with electroplating in the 1830s, it had been known that about 100,000 C per mole was needed to plate singly ionized ions. Dividing this by the number of ions per mole (that is, by Avogadro’s number), which was approximately known, the charge per ion was calculated to be about 1.6×10^{-19} C, close to the actual value.

An American physicist, Robert Millikan (1868–1953) (see [\[link\]](#)), decided to improve upon Thomson’s experiment for measuring q_e and was eventually forced to try another approach, which is now a classic experiment performed by students. The Millikan oil drop experiment is shown in [\[link\]](#).



Robert Millikan
(credit: Unknown
Author, via
Wikimedia
Commons)



The Millikan oil
drop experiment
produced the first
accurate direct
measurement of the

charge on
electrons, one of
the most
fundamental
constants in nature.

Fine drops of oil
become charged
when sprayed.
Their movement is
observed between
metal plates with a
potential applied to
oppose the
gravitational force.

The balance of
gravitational and
electric forces
allows the
calculation of the
charge on a drop.
The charge is found
to be quantized in
units of
 $-1.6 \times 10^{-19} \text{ C}$,
thus determining
directly the charge
of the excess and
missing electrons
on the oil drops.

In the Millikan oil drop experiment, fine drops of oil are sprayed from an atomizer. Some of these are charged by the process and can then be suspended between metal plates by a voltage between the plates. In this situation, the weight of the drop is balanced by the electric force:

Equation:

$$m_{\text{drop}}g = q_e E$$

The electric field is produced by the applied voltage, hence, $E = V/d$, and V is adjusted to just balance the drop's weight. The drops can be seen as points of reflected light using a microscope, but they are too small to directly measure their size and mass. The mass of the drop is determined by observing how fast it falls when the voltage is turned off. Since air resistance is very significant for these submicroscopic drops, the more massive drops fall faster than the less massive, and sophisticated sedimentation calculations can reveal their mass. Oil is used rather than water, because it does not readily evaporate, and so mass is nearly constant. Once the mass of the drop is known, the charge of the electron is given by rearranging the previous equation:

Equation:

$$q = \frac{m_{\text{drop}}g}{E} = \frac{m_{\text{drop}}gd}{V},$$

where d is the separation of the plates and V is the voltage that holds the drop motionless. (The same drop can be observed for several hours to see that it really is motionless.) By 1913 Millikan had measured the charge of the electron q_e to an accuracy of 1%, and he improved this by a factor of 10 within a few years to a value of -1.60×10^{-19} C. He also observed that all charges were multiples of the basic electron charge and that sudden changes could occur in which electrons were added or removed from the drops. For this very fundamental direct measurement of q_e and for his studies of the photoelectric effect, Millikan was awarded the 1923 Nobel Prize in Physics.

With the charge of the electron known and the charge-to-mass ratio known, the electron's mass can be calculated. It is

Equation:

$$m = \frac{q_e}{\left(\frac{q_e}{m_e}\right)}.$$

Substituting known values yields

Equation:

$$m_e = \frac{-1.60 \times 10^{-19} \text{ C}}{-1.76 \times 10^{11} \text{ C/kg}}$$

or

Equation:

$$m_e = 9.11 \times 10^{-31} \text{ kg (electron's mass),}$$

where the round-off errors have been corrected. The mass of the electron has been verified in many subsequent experiments and is now known to an accuracy of better than one part in one million. It is an incredibly small mass and remains the smallest known mass of any particle that has mass. (Some particles, such as photons, are massless and cannot be brought to rest, but travel at the speed of light.) A similar calculation gives the masses of other particles, including the proton. To three digits, the mass of the proton is now known to be

Equation:

$$m_p = 1.67 \times 10^{-27} \text{ kg (proton's mass),}$$

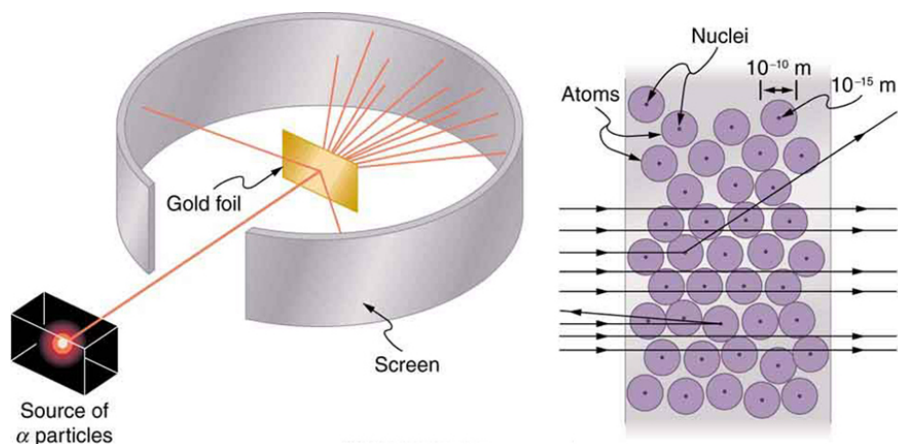
which is nearly identical to the mass of a hydrogen atom. What Thomson and Millikan had done was to prove the existence of one substructure of atoms, the electron, and further to show that it had only a tiny fraction of the mass of an atom. The nucleus of an atom contains most of its mass, and the nature of the nucleus was completely unanticipated.

Another important characteristic of quantum mechanics was also beginning to emerge. All electrons are identical to one another. The charge and mass of electrons are not average values; rather, they are unique values that all electrons have. This is true of other fundamental entities at the submicroscopic level. All protons are identical to one another, and so on.

The Nucleus

Here, we examine the first direct evidence of the size and mass of the nucleus. In later chapters, we will examine many other aspects of nuclear physics, but the basic information on nuclear size and mass is so important to understanding the atom that we consider it here.

Nuclear radioactivity was discovered in 1896, and it was soon the subject of intense study by a number of the best scientists in the world. Among them was New Zealander Lord Ernest Rutherford, who made numerous fundamental discoveries and earned the title of “father of nuclear physics.” Born in Nelson, Rutherford did his postgraduate studies at the Cavendish Laboratories in England before taking up a position at McGill University in Canada where he did the work that earned him a Nobel Prize in Chemistry in 1908. In the area of atomic and nuclear physics, there is much overlap between chemistry and physics, with physics providing the fundamental enabling theories. He returned to England in later years and had six future Nobel Prize winners as students. Rutherford used nuclear radiation to directly examine the size and mass of the atomic nucleus. The experiment he devised is shown in [\[link\]](#). A radioactive source that emits alpha radiation was placed in a lead container with a hole in one side to produce a beam of alpha particles, which are a type of ionizing radiation ejected by the nuclei of a radioactive source. A thin gold foil was placed in the beam, and the scattering of the alpha particles was observed by the glow they caused when they struck a phosphor screen.



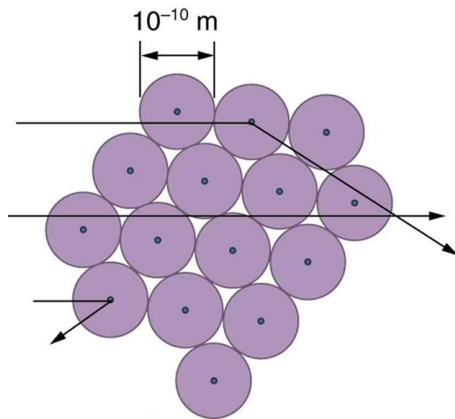
Rutherford's experiment gave direct evidence for the size and mass of the nucleus by scattering alpha particles from a thin gold foil. Alpha particles with energies of about 5 MeV are emitted from a radioactive source (which is a small metal container in which a specific amount of a radioactive material is sealed), are collimated into a beam, and fall upon the foil. The number of particles that penetrate the foil or scatter to various angles indicates that gold nuclei are very small and contain nearly all of the gold atom's mass. This is particularly indicated by the alpha particles that scatter to very large angles, much like a soccer ball bouncing off a goalie's head.

Alpha particles were known to be the doubly charged positive nuclei of helium atoms that had kinetic energies on the order of 5 MeV when emitted in nuclear decay, which is the disintegration of the nucleus of an unstable nuclide by the spontaneous emission of charged particles. These particles interact with matter mostly via the Coulomb force, and the manner in which they scatter from nuclei can reveal nuclear size and mass. This is analogous to observing how a bowling ball is scattered by an object you cannot see directly. Because the alpha particle's energy is so large compared with the typical energies associated with atoms (MeV versus eV), you would expect the alpha particles to simply crash through a thin foil much like a supersonic bowling ball would crash through a few dozen rows of bowling pins. Thomson had envisioned the atom to be a small sphere in which equal amounts of positive and negative charge were distributed evenly. The incident massive alpha particles would suffer only small deflections in such a model. Instead, Rutherford and his collaborators found that alpha particles occasionally were scattered to large angles, some even back in the direction from which they came! Detailed analysis using conservation of momentum and energy—particularly of the small number that came straight back—implied that gold nuclei are very small compared with the size of a gold atom, contain almost all of the atom's mass, and are tightly bound. Since

the gold nucleus is several times more massive than the alpha particle, a head-on collision would scatter the alpha particle straight back toward the source. In addition, the smaller the nucleus, the fewer alpha particles that would hit one head on.

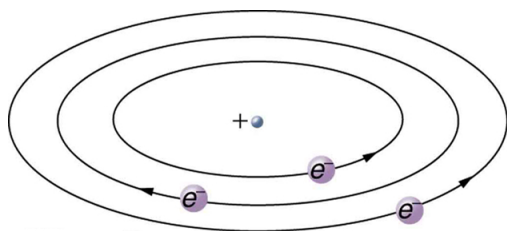
Although the results of the experiment were published by his colleagues in 1909, it took Rutherford two years to convince himself of their meaning. Like Thomson before him, Rutherford was reluctant to accept such radical results. Nature on a small scale is so unlike our classical world that even those at the forefront of discovery are sometimes surprised. Rutherford later wrote: “It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you. On consideration, I realized that this scattering backwards ... [meant] ... the greatest part of the mass of the atom was concentrated in a tiny nucleus.” In 1911, Rutherford published his analysis together with a proposed model of the atom. The size of the nucleus was determined to be about 10^{-15} m, or 100,000 times smaller than the atom. This implies a huge density, on the order of 10^{15} g/cm³, vastly unlike any macroscopic matter. Also implied is the existence of previously unknown nuclear forces to counteract the huge repulsive Coulomb forces among the positive charges in the nucleus. Huge forces would also be consistent with the large energies emitted in nuclear radiation.

The small size of the nucleus also implies that the atom is mostly empty inside. In fact, in Rutherford’s experiment, most alphas went straight through the gold foil with very little scattering, since electrons have such small masses and since the atom was mostly empty with nothing for the alpha to hit. There were already hints of this at the time Rutherford performed his experiments, since energetic electrons had been observed to penetrate thin foils more easily than expected. [\[link\]](#) shows a schematic of the atoms in a thin foil with circles representing the size of the atoms (about 10^{-10} m) and dots representing the nuclei. (The dots are not to scale—if they were, you would need a microscope to see them.) Most alpha particles miss the small nuclei and are only slightly scattered by electrons. Occasionally, (about once in 8000 times in Rutherford’s experiment), an alpha hits a nucleus head-on and is scattered straight backward.



An expanded view of the atoms in the gold foil in Rutherford's experiment. Circles represent the atoms (about 10^{-10} m in diameter), while the dots represent the nuclei (about 10^{-15} m in diameter). To be visible, the dots are much larger than scale. Most alpha particles crash through but are relatively unaffected because of their high energy and the electron's small mass. Some, however, head straight toward a nucleus and are scattered straight back. A detailed analysis gives the size and mass of the nucleus.

Based on the size and mass of the nucleus revealed by his experiment, as well as the mass of electrons, Rutherford proposed the **planetary model of the atom**. The planetary model of the atom pictures low-mass electrons orbiting a large-mass nucleus. The sizes of the electron orbits are large compared with the size of the nucleus, with mostly vacuum inside the atom. This picture is analogous to how low-mass planets in our solar system orbit the large-mass Sun at distances large compared with the size of the sun. In the atom, the attractive Coulomb force is analogous to gravitation in the planetary system. (See [\[link\]](#).) Note that a model or mental picture is needed to explain experimental results, since the atom is too small to be directly observed with visible light.



Rutherford's planetary model of the atom incorporates the characteristics of the nucleus, electrons, and the size of the atom. This model was the first to recognize the structure of atoms, in which low-mass electrons orbit a very small, massive nucleus in orbits much larger than the nucleus. The atom is mostly empty and is analogous to our planetary system.

Rutherford's planetary model of the atom was crucial to understanding the characteristics of atoms, and their interactions and energies, as we shall see in the next few sections. Also, it was an indication of how different nature is from the familiar classical world on the small, quantum mechanical scale. The discovery of a substructure to all matter in the form of atoms and molecules was now being taken a step further to reveal a substructure of atoms that was simpler than the 92 elements then known. We have continued to search for deeper substructures, such as those inside the nucleus, with some success. In later chapters, we will follow this quest in the discussion of quarks and other elementary particles, and we will look at the direction the search seems now to be heading.

Note:

PhET Explorations: Rutherford Scattering

How did Rutherford figure out the structure of the atom without being able to see it? Simulate the famous experiment in which he disproved the Plum Pudding model of the atom by observing alpha particles bouncing off atoms and determining that they must have a small core.

https://phet.colorado.edu/sims/html/rutherford-scattering/latest/rutherford-scattering_en.html

Section Summary

- Atoms are composed of negatively charged electrons, first proved to exist in cathode-ray-tube experiments, and a positively charged nucleus.
- All electrons are identical and have a charge-to-mass ratio of

Equation:

$$\frac{q_e}{m_e} = -1.76 \times 10^{11} \text{ C/kg.}$$

- The positive charge in the nuclei is carried by particles called protons, which have a charge-to-mass ratio of

Equation:

$$\frac{q_p}{m_p} = 9.57 \times 10^7 \text{ C/kg.}$$

- Mass of electron,

Equation:

$$m_e = 9.11 \times 10^{-31} \text{ kg.}$$

- Mass of proton,

Equation:

$$m_p = 1.67 \times 10^{-27} \text{ kg.}$$

- The planetary model of the atom pictures electrons orbiting the nucleus in the same way that planets orbit the sun.

Conceptual Questions

Exercise:

Problem:

What two pieces of evidence allowed the first calculation of m_e , the mass of the electron?

- The ratios q_e/m_e and q_p/m_p .
- The values of q_e and E_B .
- The ratio q_e/m_e and q_e .

Justify your response.

Exercise:

Problem:

How do the allowed orbits for electrons in atoms differ from the allowed orbits for planets around the sun? Explain how the correspondence principle applies here.

Problem Exercises**Exercise:****Problem:**

Rutherford found the size of the nucleus to be about 10^{-15} m. This implied a huge density. What would this density be for gold?

Solution:

$$6 \times 10^{20} \text{ kg/m}^3$$

Exercise:**Problem:**

In Millikan's oil-drop experiment, one looks at a small oil drop held motionless between two plates. Take the voltage between the plates to be 2033 V, and the plate separation to be 2.00 cm. The oil drop (of density 0.81 g/cm^3) has a diameter of 4.0×10^{-6} m. Find the charge on the drop, in terms of electron units.

Exercise:**Problem:**

(a) An aspiring physicist wants to build a scale model of a hydrogen atom for her science fair project. If the atom is 1.00 m in diameter, how big should she try to make the nucleus?

(b) How easy will this be to do?

Solution:

(a) $10.0\ \mu\text{m}$

(b) It isn't hard to make one of approximately this size. It would be harder to make it exactly $10.0\ \mu\text{m}$.

Glossary

cathode-ray tube

a vacuum tube containing a source of electrons and a screen to view images

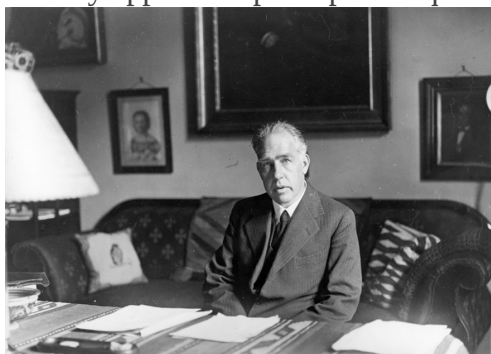
planetary model of the atom

the most familiar model or illustration of the structure of the atom

Bohr's Theory of the Hydrogen Atom

- Describe the mysteries of atomic spectra.
- Explain Bohr's theory of the hydrogen atom.
- Explain Bohr's planetary model of the atom.
- Illustrate energy state using the energy-level diagram.
- Describe the triumphs and limits of Bohr's theory.

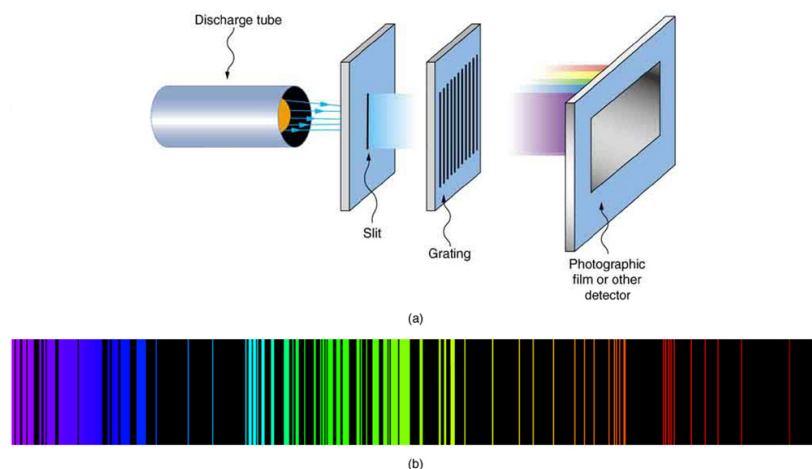
The great Danish physicist Niels Bohr (1885–1962) made immediate use of Rutherford's planetary model of the atom. ([link](#)). Bohr became convinced of its validity and spent part of 1912 at Rutherford's laboratory. In 1913, after returning to Copenhagen, he began publishing his theory of the simplest atom, hydrogen, based on the planetary model of the atom. For decades, many questions had been asked about atomic characteristics. From their sizes to their spectra, much was known about atoms, but little had been explained in terms of the laws of physics. Bohr's theory explained the atomic spectrum of hydrogen and established new and broadly applicable principles in quantum mechanics.



Niels Bohr, Danish physicist, used the planetary model of the atom to explain the atomic spectrum and size of the hydrogen atom. His many contributions to the development of atomic physics and quantum mechanics, his personal influence on many students and colleagues, and his personal integrity, especially in the face of Nazi oppression, earned him a prominent place in history. (credit: Unknown Author, via Wikimedia Commons)

Mysteries of Atomic Spectra

As noted in [Quantization of Energy](#), the energies of some small systems are quantized. Atomic and molecular emission and absorption spectra have been known for over a century to be discrete (or quantized). (See [\[link\]](#).) Maxwell and others had realized that there must be a connection between the spectrum of an atom and its structure, something like the resonant frequencies of musical instruments. But, in spite of years of efforts by many great minds, no one had a workable theory. (It was a running joke that any theory of atomic and molecular spectra could be destroyed by throwing a book of data at it, so complex were the spectra.) Following Einstein's proposal of photons with quantized energies directly proportional to their wavelengths, it became even more evident that electrons in atoms can exist only in discrete orbits.



Part (a) shows, from left to right, a discharge tube, slit, and diffraction grating producing a line spectrum. Part (b) shows the emission line spectrum for iron. The discrete lines imply quantized energy states for the atoms that produce them. The line spectrum for each element is unique, providing a powerful and much used analytical tool, and many line spectra were well known for many years before they could be explained with physics.
(credit for (b): Yttrium91, Wikimedia Commons)

In some cases, it had been possible to devise formulas that described the emission spectra. As you might expect, the simplest atom—hydrogen, with its single electron—has a relatively simple spectrum. The hydrogen spectrum had been observed in the infrared (IR), visible, and ultraviolet (UV), and several series of spectral lines had been observed. (See [\[link\]](#).) These series are named after early researchers who studied them in particular depth.

The observed **hydrogen-spectrum wavelengths** can be calculated using the following formula:

Equation:

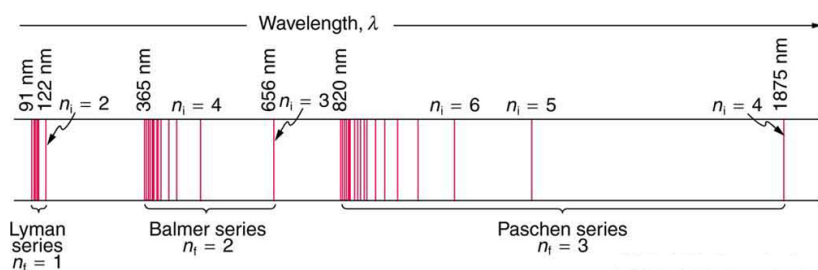
$$\frac{1}{\lambda} = R \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right),$$

where λ is the wavelength of the emitted EM radiation and R is the **Rydberg constant**, determined by the experiment to be

Equation:

$$R = 1.097 \times 10^7 / \text{m (or m}^{-1}\text{)}.$$

The constant n_f is a positive integer associated with a specific series. For the Lyman series, $n_f = 1$; for the Balmer series, $n_f = 2$; for the Paschen series, $n_f = 3$; and so on. The Lyman series is entirely in the UV, while part of the Balmer series is visible with the remainder UV. The Paschen series and all the rest are entirely IR. There are apparently an unlimited number of series, although they lie progressively farther into the infrared and become difficult to observe as n_f increases. The constant n_i is a positive integer, but it must be greater than n_f . Thus, for the Balmer series, $n_f = 2$ and $n_i = 3, 4, 5, 6, \dots$. Note that n_i can approach infinity. While the formula in the wavelengths equation was just a recipe designed to fit data and was not based on physical principles, it did imply a deeper meaning. Balmer first devised the formula for his series alone, and it was later found to describe all the other series by using different values of n_f . Bohr was the first to comprehend the deeper meaning. Again, we see the interplay between experiment and theory in physics. Experimentally, the spectra were well established, an equation was found to fit the experimental data, but the theoretical foundation was missing.



A schematic of the hydrogen spectrum shows several series named for those who contributed most to their determination. Part of the Balmer series is in the visible spectrum, while the Lyman series is entirely in the UV, and the Paschen series and others are in the IR. Values of n_f and n_i are shown for some of the lines.

Example:**Calculating Wave Interference of a Hydrogen Line**

What is the distance between the slits of a grating that produces a first-order maximum for the second Balmer line at an angle of 15° ?

Strategy and Concept

For an Integrated Concept problem, we must first identify the physical principles involved. In this example, we need to know (a) the wavelength of light as well as (b) conditions for an interference maximum for the pattern from a double slit. Part (a) deals with a topic of the present chapter, while part (b) considers the wave interference material of [Wave Optics](#).

Solution for (a)

Hydrogen spectrum wavelength. The Balmer series requires that $n_f = 2$. The first line in the series is taken to be for $n_i = 3$, and so the second would have $n_i = 4$.

The calculation is a straightforward application of the wavelength equation. Entering the determined values for n_f and n_i yields

Equation:

$$\begin{aligned}\frac{1}{\lambda} &= R\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \\ &= (1.097 \times 10^7 \text{ m}^{-1})\left(\frac{1}{2^2} - \frac{1}{4^2}\right) \\ &= 2.057 \times 10^6 \text{ m}^{-1}.\end{aligned}$$

Inverting to find λ gives

Equation:

$$\begin{aligned}\lambda &= \frac{1}{2.057 \times 10^6 \text{ m}^{-1}} = 486 \times 10^{-9} \text{ m} \\ &= 486 \text{ nm}.\end{aligned}$$

Discussion for (a)

This is indeed the experimentally observed wavelength, corresponding to the second (blue-green) line in the Balmer series. More impressive is the fact that the same simple recipe predicts *all* of the hydrogen spectrum lines, including new ones observed in subsequent experiments. What is nature telling us?

Solution for (b)

Double-slit interference ([Wave Optics](#)). To obtain constructive interference for a double slit, the path length difference from two slits must be an integral multiple of the wavelength. This condition was expressed by the equation

Equation:

$$d \sin \theta = m\lambda,$$

where d is the distance between slits and θ is the angle from the original direction of the beam. The number m is the order of the interference; $m = 1$ in this example. Solving for d and entering known values yields

Equation:

$$d = \frac{(1)(486 \text{ nm})}{\sin 15^\circ} = 1.88 \times 10^{-6} \text{ m}.$$

Discussion for (b)

This number is similar to those used in the interference examples of [Introduction to Quantum Physics](#) (and is close to the spacing between slits in commonly used diffraction glasses).

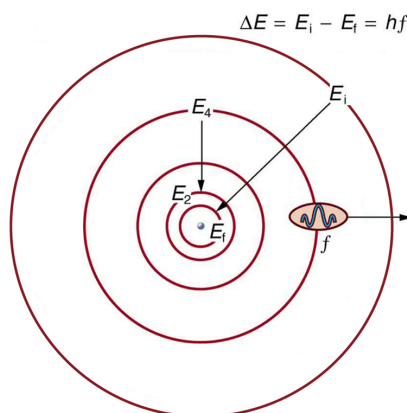
Bohr's Solution for Hydrogen

Bohr was able to derive the formula for the hydrogen spectrum using basic physics, the planetary model of the atom, and some very important new proposals. His first proposal is that only certain orbits are allowed: we say that *the orbits of electrons in atoms are quantized*. Each orbit has a different energy, and electrons can move to a higher orbit by absorbing energy and drop to a lower orbit by emitting energy. If the orbits are quantized, the amount of energy absorbed or emitted is also quantized, producing discrete spectra. Photon absorption and emission are among the primary methods of transferring energy into and out of atoms. The energies of the photons are quantized, and their energy is explained as being equal to the change in energy of the electron when it moves from one orbit to another. In equation form, this is

Equation:

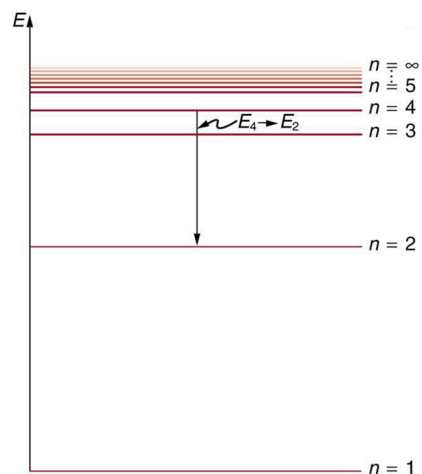
$$\Delta E = hf = E_i - E_f.$$

Here, ΔE is the change in energy between the initial and final orbits, and hf is the energy of the absorbed or emitted photon. It is quite logical (that is, expected from our everyday experience) that energy is involved in changing orbits. A blast of energy is required for the space shuttle, for example, to climb to a higher orbit. What is not expected is that atomic orbits should be quantized. This is not observed for satellites or planets, which can have any orbit given the proper energy. (See [\[link\]](#).)



The planetary model of the atom, as modified by Bohr, has the orbits of the electrons quantized. Only certain orbits are allowed, explaining why atomic spectra are discrete (quantized). The energy carried away from an atom by a photon comes from the electron dropping from one allowed orbit to another and is thus quantized. This is likewise true for atomic absorption of photons.

[\[link\]](#) shows an **energy-level diagram**, a convenient way to display energy states. In the present discussion, we take these to be the allowed energy levels of the electron. Energy is plotted vertically with the lowest or ground state at the bottom and with excited states above. Given the energies of the lines in an atomic spectrum, it is possible (although sometimes very difficult) to determine the energy levels of an atom. Energy-level diagrams are used for many systems, including molecules and nuclei. A theory of the atom or any other system must predict its energies based on the physics of the system.



An energy-level diagram plots energy vertically and is useful in visualizing the energy states of a system and the transitions between them.

This diagram is for the hydrogen-atom electrons, showing a transition between two orbits having energies E_4 and E_2 .

Bohr was clever enough to find a way to calculate the electron orbital energies in hydrogen. This was an important first step that has been improved upon, but it is well worth repeating here, because it does correctly describe many characteristics of hydrogen. Assuming circular orbits, Bohr proposed that the **angular momentum L of an electron in its orbit is quantized**, that is, it has only specific, discrete values. The value for L is given by the formula

Equation:

$$L = m_e v r_n = n \frac{h}{2\pi} (n = 1, 2, 3, \dots),$$

where L is the angular momentum, m_e is the electron's mass, r_n is the radius of the n th orbit, and h is Planck's constant. Note that angular momentum is $L = I\omega$. For a small object at a radius r , $I = mr^2$ and $\omega = v/r$, so that $L = (mr^2)(v/r) = mvr$. Quantization says that this value of mvr can only be equal to $h/2$, $2h/2$, $3h/2$, etc. At the time, Bohr himself did not know why angular momentum should be quantized, but using this assumption he was

able to calculate the energies in the hydrogen spectrum, something no one else had done at the time.

From Bohr's assumptions, we will now derive a number of important properties of the hydrogen atom from the classical physics we have covered in the text. We start by noting the centripetal force causing the electron to follow a circular path is supplied by the Coulomb force. To be more general, we note that this analysis is valid for any single-electron atom. So, if a nucleus has Z protons ($Z = 1$ for hydrogen, 2 for helium, etc.) and only one electron, that atom is called a **hydrogen-like atom**. The spectra of hydrogen-like ions are similar to hydrogen, but shifted to higher energy by the greater attractive force between the electron and nucleus. The magnitude of the centripetal force is $m_e v^2 / r_n$, while the Coulomb force is $k(Zq_e)(q_e)/r_n^2$. The tacit assumption here is that the nucleus is more massive than the stationary electron, and the electron orbits about it. This is consistent with the planetary model of the atom. Equating these,

Equation:

$$k \frac{Zq_e^2}{r_n^2} = \frac{m_e v^2}{r_n} \text{ (Coulomb = centripetal).}$$

Angular momentum quantization is stated in an earlier equation. We solve that equation for v , substitute it into the above, and rearrange the expression to obtain the radius of the orbit. This yields:

Equation:

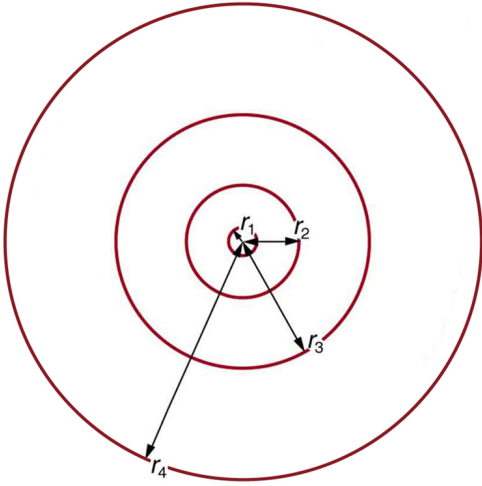
$$r_n = \frac{n^2}{Z} a_B, \text{ for allowed orbits } (n = 1, 2, 3, \dots),$$

where a_B is defined to be the **Bohr radius**, since for the lowest orbit ($n = 1$) and for hydrogen ($Z = 1$), $r_1 = a_B$. It is left for this chapter's Problems and Exercises to show that the Bohr radius is

Equation:

$$a_B = \frac{h^2}{4\pi^2 m_e k q_e^2} = 0.529 \times 10^{-10} \text{ m.}$$

These last two equations can be used to calculate the **radii of the allowed (quantized) electron orbits in any hydrogen-like atom**. It is impressive that the formula gives the correct size of hydrogen, which is measured experimentally to be very close to the Bohr radius. The earlier equation also tells us that the orbital radius is proportional to n^2 , as illustrated in [\[link\]](#).



The allowed electron orbits in hydrogen have the radii shown. These radii were first calculated by Bohr and are given by the equation $r_n = \frac{n^2}{Z} a_B$. The lowest orbit has the experimentally verified diameter of a hydrogen atom.

To get the electron orbital energies, we start by noting that the electron energy is the sum of its kinetic and potential energy:

Equation:

$$E_n = \text{KE} + \text{PE}.$$

Kinetic energy is the familiar $\text{KE} = (1/2)m_e v^2$, assuming the electron is not moving at relativistic speeds. Potential energy for the electron is electrical, or $\text{PE} = q_e V$, where V is the potential due to the nucleus, which looks like a point charge. The nucleus has a positive charge Zq_e ; thus, $V = kZq_e/r_n$, recalling an earlier equation for the potential due to a point charge. Since the electron's charge is negative, we see that $\text{PE} = -kZq_e/r_n$. Entering the expressions for KE and PE, we find

Equation:

$$E_n = \frac{1}{2}m_e v^2 - k \frac{Zq_e^2}{r_n}.$$

Now we substitute r_n and v from earlier equations into the above expression for energy. Algebraic manipulation yields

Equation:

$$E_n = -\frac{Z^2}{n^2} E_0 (n = 1, 2, 3, \dots)$$

for the orbital **energies of hydrogen-like atoms**. Here, E_0 is the **ground-state energy** ($n = 1$) for hydrogen ($Z = 1$) and is given by

Equation:

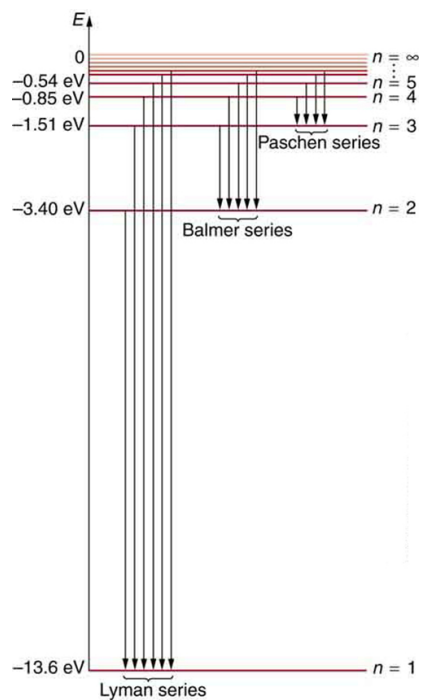
$$E_0 = \frac{2\pi^2 q_e^4 m_e k^2}{h^2} = 13.6 \text{ eV}.$$

Thus, for hydrogen,

Equation:

$$E_n = -\frac{13.6 \text{ eV}}{n^2} (n = 1, 2, 3, \dots).$$

[\[link\]](#) shows an energy-level diagram for hydrogen that also illustrates how the various spectral series for hydrogen are related to transitions between energy levels.



Energy-level diagram for

hydrogen showing the Lyman, Balmer, and Paschen series of transitions. The orbital energies are calculated using the above equation, first derived by Bohr.

Electron total energies are negative, since the electron is bound to the nucleus, analogous to being in a hole without enough kinetic energy to escape. As n approaches infinity, the total energy becomes zero. This corresponds to a free electron with no kinetic energy, since r_n gets very large for large n , and the electric potential energy thus becomes zero. Thus, 13.6 eV is needed to ionize hydrogen (to go from -13.6 eV to 0, or unbound), an experimentally verified number. Given more energy, the electron becomes unbound with some kinetic energy. For example, giving 15.0 eV to an electron in the ground state of hydrogen strips it from the atom and leaves it with 1.4 eV of kinetic energy.

Finally, let us consider the energy of a photon emitted in a downward transition, given by the equation to be

Equation:

$$\Delta E = hf = E_i - E_f.$$

Substituting $E_n = (-13.6 \text{ eV}/n^2)$, we see that

Equation:

$$hf = (13.6 \text{ eV}) \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right).$$

Dividing both sides of this equation by hc gives an expression for $1/\lambda$:

Equation:

$$\frac{hf}{hc} = \frac{f}{c} = \frac{1}{\lambda} = \frac{(13.6 \text{ eV})}{hc} \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right).$$

It can be shown that

Equation:

$$\left(\frac{13.6 \text{ eV}}{hc} \right) = \frac{(13.6 \text{ eV})(1.602 \times 10^{-19} \text{ J/eV})}{(6.626 \times 10^{-34} \text{ J}\cdot\text{s})(2.998 \times 10^8 \text{ m/s})} = 1.097 \times 10^7 \text{ m}^{-1} = R$$

is the **Rydberg constant**. Thus, we have used Bohr's assumptions to derive the formula first proposed by Balmer years earlier as a recipe to fit experimental data.

Equation:

$$\frac{1}{\lambda} = R \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right)$$

We see that Bohr's theory of the hydrogen atom answers the question as to why this previously known formula describes the hydrogen spectrum. It is because the energy levels are proportional to $1/n^2$, where n is a non-negative integer. A downward transition releases energy, and so n_i must be greater than n_f . The various series are those where the transitions end on a certain level. For the Lyman series, $n_f = 1$ — that is, all the transitions end in the ground state (see also [link](#)). For the Balmer series, $n_f = 2$, or all the transitions end in the first excited state; and so on. What was once a recipe is now based in physics, and something new is emerging—angular momentum is quantized.

Triumphs and Limits of the Bohr Theory

Bohr did what no one had been able to do before. Not only did he explain the spectrum of hydrogen, he correctly calculated the size of the atom from basic physics. Some of his ideas are broadly applicable. Electron orbital energies are quantized in all atoms and molecules. Angular momentum is quantized. The electrons do not spiral into the nucleus, as expected classically (accelerated charges radiate, so that the electron orbits classically would decay quickly, and the electrons would sit on the nucleus—matter would collapse). These are major triumphs.

But there are limits to Bohr's theory. It cannot be applied to multielectron atoms, even one as simple as a two-electron helium atom. Bohr's model is what we call *semiclassical*. The orbits are quantized (nonclassical) but are assumed to be simple circular paths (classical). As quantum mechanics was developed, it became clear that there are no well-defined orbits; rather, there are clouds of probability. Bohr's theory also did not explain that some spectral lines are doublets (split into two) when examined closely. We shall examine many of these aspects of quantum mechanics in more detail, but it should be kept in mind that Bohr did not fail. Rather, he made very important steps along the path to greater knowledge and laid the foundation for all of atomic physics that has since evolved.

Note:

PhET Explorations: Models of the Hydrogen Atom

How did scientists figure out the structure of atoms without looking at them? Try out different models by shooting light at the atom. Check how the prediction of the model

matches the experimental results.

<https://archive.cnx.org/specials/d77cc1d0-33e4-11e6-b016-6726afecd2be/hydrogen-atom/#sim-hydrogen-atom>

Section Summary

- The planetary model of the atom pictures electrons orbiting the nucleus in the way that planets orbit the sun. Bohr used the planetary model to develop the first reasonable theory of hydrogen, the simplest atom. Atomic and molecular spectra are quantized, with hydrogen spectrum wavelengths given by the formula

Equation:

$$\frac{1}{\lambda} = R \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right),$$

where λ is the wavelength of the emitted EM radiation and R is the Rydberg constant, which has the value

Equation:

$$R = 1.097 \times 10^7 \text{ m}^{-1}.$$

- The constants n_i and n_f are positive integers, and n_i must be greater than n_f .
- Bohr correctly proposed that the energy and radii of the orbits of electrons in atoms are quantized, with energy for transitions between orbits given by

Equation:

$$\Delta E = hf = E_i - E_f,$$

where ΔE is the change in energy between the initial and final orbits and hf is the energy of an absorbed or emitted photon. It is useful to plot orbital energies on a vertical graph called an energy-level diagram.

- Bohr proposed that the allowed orbits are circular and must have quantized orbital angular momentum given by

Equation:

$$L = m_e v r_n = n \frac{h}{2\pi} (n = 1, 2, 3 \dots),$$

where L is the angular momentum, r_n is the radius of the n th orbit, and h is Planck's constant. For all one-electron (hydrogen-like) atoms, the radius of an orbit is given by

Equation:

$$r_n = \frac{n^2}{Z} a_B (\text{allowed orbits } n = 1, 2, 3, \dots),$$

Z is the atomic number of an element (the number of electrons it has when neutral) and a_B is defined to be the Bohr radius, which is

Equation:

$$a_B = \frac{h^2}{4\pi^2 m_e k q_e^2} = 0.529 \times 10^{-10} \text{ m}.$$

- Furthermore, the energies of hydrogen-like atoms are given by
Equation:

$$E_n = -\frac{Z^2}{n^2} E_0 (n = 1, 2, 3 \dots),$$

where E_0 is the ground-state energy and is given by

Equation:

$$E_0 = \frac{2\pi^2 q_e^4 m_e k^2}{h^2} = 13.6 \text{ eV}.$$

Thus, for hydrogen,

Equation:

$$E_n = -\frac{13.6 \text{ eV}}{n^2} (n = 1, 2, 3 \dots).$$

- The Bohr Theory gives accurate values for the energy levels in hydrogen-like atoms, but it has been improved upon in several respects.

Conceptual Questions

Exercise:

Problem:

How do the allowed orbits for electrons in atoms differ from the allowed orbits for planets around the sun? Explain how the correspondence principle applies here.

Exercise:

Problem:

Explain how Bohr's rule for the quantization of electron orbital angular momentum differs from the actual rule.

Exercise:

Problem:

What is a hydrogen-like atom, and how are the energies and radii of its electron orbits related to those in hydrogen?

Problems & Exercises**Exercise:****Problem:**

By calculating its wavelength, show that the first line in the Lyman series is UV radiation.

Solution:

$$\frac{1}{\lambda} = R \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right) \Rightarrow \lambda = \frac{1}{R} \left[\frac{(n_i \cdot n_f)^2}{n_i^2 - n_f^2} \right]; n_i = 2, n_f = 1, \text{ so that}$$

$$\lambda = \left(\frac{\text{m}}{1.097 \times 10^7} \right) \left[\frac{(2 \times 1)^2}{2^2 - 1^2} \right] = 1.22 \times 10^{-7} \text{ m} = 122 \text{ nm}, \text{ which is UV radiation.}$$

Exercise:**Problem:**

Find the wavelength of the third line in the Lyman series, and identify the type of EM radiation.

Exercise:**Problem:**

Look up the values of the quantities in $a_B = \frac{h^2}{4\pi^2 m_e k q_e^2}$, and verify that the Bohr radius a_B is $0.529 \times 10^{-10} \text{ m}$.

Solution:

$$a_B = \frac{h^2}{4\pi^2 m_e k q_e^2} = \frac{(6.626 \times 10^{-34} \text{ J}\cdot\text{s})^2}{4\pi^2 (9.109 \times 10^{-31} \text{ kg})(8.988 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2)(1)(1.602 \times 10^{-19} \text{ C})^2} = 0.529 \times 10^{-10} \text{ m}$$

Exercise:

Problem: Verify that the ground state energy E_0 is 13.6 eV by using $E_0 = \frac{2\pi^2 q_e^4 m_e k^2}{h^2}$.

Exercise:

Problem:

If a hydrogen atom has its electron in the $n = 4$ state, how much energy in eV is needed to ionize it?

Solution:

0.850 eV

Exercise:**Problem:**

A hydrogen atom in an excited state can be ionized with less energy than when it is in its ground state. What is n for a hydrogen atom if 0.850 eV of energy can ionize it?

Exercise:**Problem:**

Find the radius of a hydrogen atom in the $n = 2$ state according to Bohr's theory.

Solution:

$2.12 \times 10^{-10} \text{ m}$

Exercise:**Problem:**

Show that $(13.6 \text{ eV})/hc = 1.097 \times 10^7 \text{ m} = R$ (Rydberg's constant), as discussed in the text.

Exercise:**Problem:**

What is the smallest-wavelength line in the Balmer series? Is it in the visible part of the spectrum?

Solution:

365 nm

It is in the ultraviolet.

Exercise:**Problem:**

Show that the entire Paschen series is in the infrared part of the spectrum. To do this, you only need to calculate the shortest wavelength in the series.

Exercise:

Problem:

Do the Balmer and Lyman series overlap? To answer this, calculate the shortest-wavelength Balmer line and the longest-wavelength Lyman line.

Solution:

No overlap

365 nm

122 nm

Exercise:

Problem:

(a) Which line in the Balmer series is the first one in the UV part of the spectrum?

(b) How many Balmer series lines are in the visible part of the spectrum?

(c) How many are in the UV?

Exercise:

Problem:

A wavelength of $4.653\text{ }\mu\text{m}$ is observed in a hydrogen spectrum for a transition that ends in the $n_f = 5$ level. What was n_i for the initial level of the electron?

Solution:

7

Exercise:

Problem:

A singly ionized helium ion has only one electron and is denoted He^+ . What is the ion's radius in the ground state compared to the Bohr radius of hydrogen atom?

Exercise:

Problem:

A beryllium ion with a single electron (denoted Be^{3+}) is in an excited state with radius the same as that of the ground state of hydrogen.

(a) What is n for the Be^{3+} ion?

(b) How much energy in eV is needed to ionize the ion from this excited state?

Solution:

(a) 2

(b) 54.4 eV

Exercise:

Problem:

Atoms can be ionized by thermal collisions, such as at the high temperatures found in the solar corona. One such ion is C^{+5} , a carbon atom with only a single electron.

(a) By what factor are the energies of its hydrogen-like levels greater than those of hydrogen?

(b) What is the wavelength of the first line in this ion's Paschen series?

(c) What type of EM radiation is this?

Exercise:

Problem:

Verify Equations $r_n = \frac{n^2}{Z} a_B$ and $a_B = \frac{h^2}{4\pi^2 m_e k q_e^2} = 0.529 \times 10^{-10} \text{ m}$ using the approach stated in the text. That is, equate the Coulomb and centripetal forces and then insert an expression for velocity from the condition for angular momentum quantization.

Solution:

$\frac{kZq_e^2}{r_n^2} = \frac{m_e V^2}{r_n}$, so that $r_n = \frac{kZq_e^2}{m_e V^2} = \frac{kZq_e^2}{m_e} \frac{1}{V^2}$. From the equation $m_e v r_n = n \frac{h}{2\pi}$, we can substitute for the velocity, giving: $r_n = \frac{kZq_e^2}{m_e} \cdot \frac{4\pi^2 m_e^2 r_n^2}{n^2 h^2}$ so that $r_n = \frac{n^2}{Z} \frac{h^2}{4\pi^2 m_e k q_e^2} = \frac{n^2}{Z} a_B$, where $a_B = \frac{h^2}{4\pi^2 m_e k q_e^2}$.

Exercise:

Problem:

The wavelength of the four Balmer series lines for hydrogen are found to be 410.3, 434.2, 486.3, and 656.5 nm. What average percentage difference is found between these wavelength numbers and those predicted by $\frac{1}{\lambda} = R \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right)$? It is amazing how well a simple formula (disconnected originally from theory) could duplicate this phenomenon.

Glossary

hydrogen spectrum wavelengths

the wavelengths of visible light from hydrogen; can be calculated by

$$\frac{1}{\lambda} = R \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right)$$

Rydberg constant

a physical constant related to the atomic spectra with an established value of

$$1.097 \times 10^7 \text{ m}^{-1}$$

double-slit interference

an experiment in which waves or particles from a single source impinge upon two slits so that the resulting interference pattern may be observed

energy-level diagram

a diagram used to analyze the energy level of electrons in the orbits of an atom

Bohr radius

the mean radius of the orbit of an electron around the nucleus of a hydrogen atom in its ground state

hydrogen-like atom

any atom with only a single electron

energies of hydrogen-like atoms

Bohr formula for energies of electron states in hydrogen-like atoms:

$$E_n = -\frac{Z^2}{n^2} E_0 (n = 1, 2, 3, \dots)$$

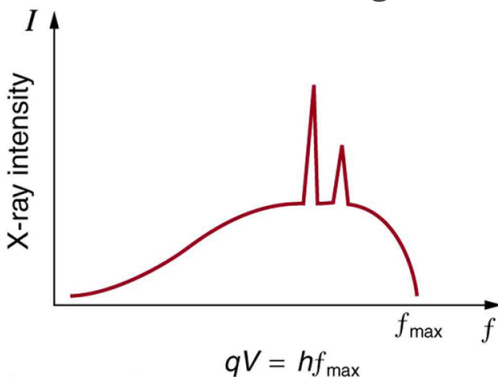
X Rays: Atomic Origins and Applications

- Define x-ray tube and its spectrum.
- Show the x-ray characteristic energy.
- Specify the use of x rays in medical observations.
- Explain the use of x rays in CT scanners in diagnostics.

Each type of atom (or element) has its own characteristic electromagnetic spectrum. **X rays** lie at the high-frequency end of an atom's spectrum and are characteristic of the atom as well. In this section, we explore characteristic x rays and some of their important applications.

We have previously discussed x rays as a part of the electromagnetic spectrum in [Photon Energies and the Electromagnetic Spectrum](#). That module illustrated how an x-ray tube (a specialized CRT) produces x rays. Electrons emitted from a hot filament are accelerated with a high voltage, gaining significant kinetic energy and striking the anode.

There are two processes by which x rays are produced in the anode of an x-ray tube. In one process, the deceleration of electrons produces x rays, and these x rays are called *bremsstrahlung*, or braking radiation. The second process is atomic in nature and produces *characteristic x rays*, so called because they are characteristic of the anode material. The x-ray spectrum in [\[link\]](#) is typical of what is produced by an x-ray tube, showing a broad curve of bremsstrahlung radiation with characteristic x-ray peaks on it.



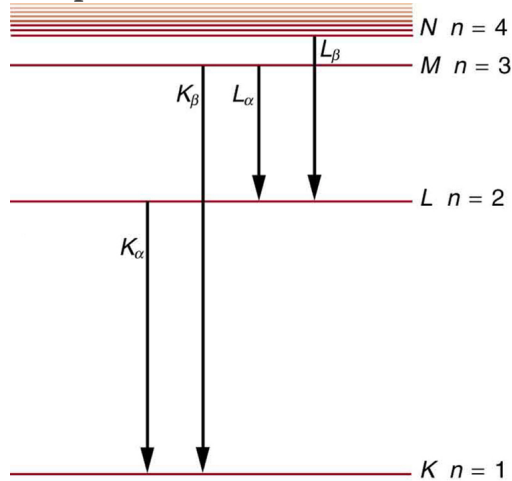
X-ray spectrum obtained when energetic electrons strike a material, such as

in the anode of a CRT.
The smooth part of the
spectrum is
bremsstrahlung radiation,
while the peaks are
characteristic of the
anode material. A
different anode material
would have characteristic
x-ray peaks at different
frequencies.

The spectrum in [\[link\]](#) is collected over a period of time in which many electrons strike the anode, with a variety of possible outcomes for each hit. The broad range of x-ray energies in the bremsstrahlung radiation indicates that an incident electron's energy is not usually converted entirely into photon energy. The highest-energy x ray produced is one for which all of the electron's energy was converted to photon energy. Thus the accelerating voltage and the maximum x-ray energy are related by conservation of energy. Electric potential energy is converted to kinetic energy and then to photon energy, so that $E_{\text{max}} = hf_{\text{max}} = q_e V$. Units of electron volts are convenient. For example, a 100-kV accelerating voltage produces x-ray photons with a maximum energy of 100 keV.

Some electrons excite atoms in the anode. Part of the energy that they deposit by collision with an atom results in one or more of the atom's inner electrons being knocked into a higher orbit or the atom being ionized. When the anode's atoms de-excite, they emit characteristic electromagnetic radiation. The most energetic of these are produced when an inner-shell vacancy is filled—that is, when an $n = 1$ or $n = 2$ shell electron has been excited to a higher level, and another electron falls into the vacant spot. A *characteristic x ray* (see [Photon Energies and the Electromagnetic Spectrum](#)) is electromagnetic (EM) radiation emitted by an atom when an inner-shell vacancy is filled. [\[link\]](#) shows a representative energy-level diagram that illustrates the labeling of characteristic x rays. X rays created

when an electron falls into an $n = 1$ shell vacancy are called K_α when they come from the next higher level; that is, an $n = 2$ to $n = 1$ transition. The labels K, L, M, \dots come from the older alphabetical labeling of shells starting with K rather than using the principal quantum numbers 1, 2, 3, A more energetic K_β x ray is produced when an electron falls into an $n = 1$ shell vacancy from the $n = 3$ shell; that is, an $n = 3$ to $n = 1$ transition. Similarly, when an electron falls into the $n = 2$ shell from the $n = 3$ shell, an L_α x ray is created. The energies of these x rays depend on the energies of electron states in the particular atom and, thus, are characteristic of that element: every element has its own set of x-ray energies. This property can be used to identify elements, for example, to find trace (small) amounts of an element in an environmental or biological sample.



A characteristic x ray is emitted when an electron fills an inner-shell vacancy, as shown for several transitions in this approximate energy level diagram for a multiple-electron atom.

Characteristic x rays are labeled according to the shell that had the vacancy and the shell from which

the electron came. A K_α x ray, for example, is produced when an electron coming from the $n = 2$ shell fills the $n = 1$ shell vacancy.

Example:

Characteristic X-Ray Energy

Calculate the approximate energy of a K_α x ray from a tungsten anode in an x-ray tube.

Strategy

How do we calculate energies in a multiple-electron atom? In the case of characteristic x rays, the following approximate calculation is reasonable. Characteristic x rays are produced when an inner-shell vacancy is filled. Inner-shell electrons are nearer the nucleus than others in an atom and thus feel little net effect from the others. This is similar to what happens inside a charged conductor, where its excess charge is distributed over the surface so that it produces no electric field inside. It is reasonable to assume the inner-shell electrons have hydrogen-like energies, as given by

$E_n = -\frac{Z^2}{n^2} E_0$ ($n = 1, 2, 3, \dots$). As noted, a K_α x ray is produced by an $n = 2$ to $n = 1$ transition. Since there are two electrons in a filled K shell, a vacancy would leave one electron, so that the effective charge would be $Z - 1$ rather than Z . For tungsten, $Z = 74$, so that the effective charge is 73.

Solution

$E_n = -\frac{Z^2}{n^2} E_0$ ($n = 1, 2, 3, \dots$) gives the orbital energies for hydrogen-like atoms to be $E_n = -(Z^2/n^2)E_0$, where $E_0 = 13.6$ eV. As noted, the effective Z is 73. Now the K_α x-ray energy is given by

Equation:

$$E_{K_\alpha} = \Delta E = E_i - E_f = E_2 - E_1,$$

where

Equation:

$$E_1 = -\frac{Z^2}{1^2} E_0 = -\frac{73^2}{1} (13.6 \text{ eV}) = -72.5 \text{ keV}$$

and

Equation:

$$E_2 = -\frac{Z^2}{2^2} E_0 = -\frac{73^2}{4} (13.6 \text{ eV}) = -18.1 \text{ keV}.$$

Thus,

Equation:

$$E_{K_\alpha} = -18.1 \text{ keV} - (-72.5 \text{ keV}) = 54.4 \text{ keV}.$$

Discussion

This large photon energy is typical of characteristic x rays from heavy elements. It is large compared with other atomic emissions because it is produced when an inner-shell vacancy is filled, and inner-shell electrons are tightly bound. Characteristic x ray energies become progressively larger for heavier elements because their energy increases approximately as Z^2 . Significant accelerating voltage is needed to create these inner-shell vacancies. In the case of tungsten, at least 72.5 kV is needed, because other shells are filled and you cannot simply bump one electron to a higher filled shell. Tungsten is a common anode material in x-ray tubes; so much of the energy of the impinging electrons is absorbed, raising its temperature, that a high-melting-point material like tungsten is required.

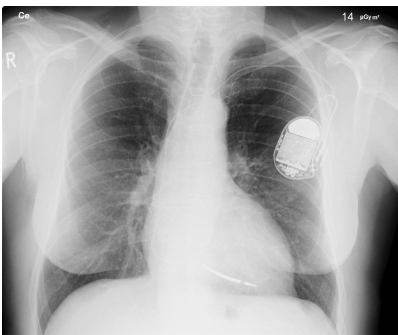
Medical and Other Diagnostic Uses of X-rays

All of us can identify diagnostic uses of x-ray photons. Among these are the universal dental and medical x rays that have become an essential part of medical diagnostics. (See [\[link\]](#) and [\[link\]](#).) X rays are also used to inspect

our luggage at airports, as shown in [\[link\]](#), and for early detection of cracks in crucial aircraft components. An x ray is not only a noun meaning high-energy photon, it also is an image produced by x rays, and it has been made into a familiar verb—to be x-rayed.

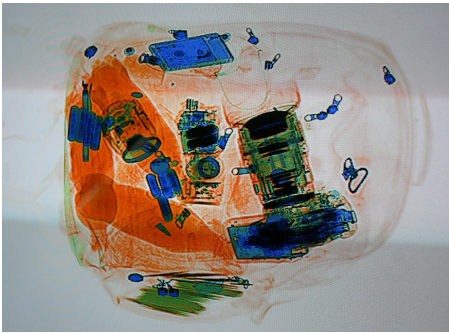


An x-ray image reveals
fillings in a person's
teeth. (credit: Dmitry G,
Wikimedia Commons)



This x-ray image of
a person's chest
shows many
details, including
an artificial
pacemaker. (credit:
Sunzi99,

Wikimedia
Commons)



This x-ray image
shows the contents of
a piece of luggage.

The denser the
material, the darker
the shadow. (credit:
IDuke, Wikimedia
Commons)

The most common x-ray images are simple shadows. Since x-ray photons have high energies, they penetrate materials that are opaque to visible light. The more energy an x-ray photon has, the more material it will penetrate. So an x-ray tube may be operated at 50.0 kV for a chest x ray, whereas it may need to be operated at 100 kV to examine a broken leg in a cast. The depth of penetration is related to the density of the material as well as to the energy of the photon. The denser the material, the fewer x-ray photons get through and the darker the shadow. Thus x rays excel at detecting breaks in bones and in imaging other physiological structures, such as some tumors, that differ in density from surrounding material. Because of their high photon energy, x rays produce significant ionization in materials and

damage cells in biological organisms. Modern uses minimize exposure to the patient and eliminate exposure to others. Biological effects of x rays will be explored in the next chapter along with other types of ionizing radiation such as those produced by nuclei.

As the x-ray energy increases, the Compton effect (see [Photon Momentum](#)) becomes more important in the attenuation of the x rays. Here, the x ray scatters from an outer electron shell of the atom, giving the ejected electron some kinetic energy while losing energy itself. The probability for attenuation of the x rays depends upon the number of electrons present (the material's density) as well as the thickness of the material. Chemical composition of the medium, as characterized by its atomic number Z , is not important here. Low-energy x rays provide better contrast (sharper images). However, due to greater attenuation and less scattering, they are more absorbed by thicker materials. Greater contrast can be achieved by injecting a substance with a large atomic number, such as barium or iodine. The structure of the part of the body that contains the substance (e.g., the gastrointestinal tract or the abdomen) can easily be seen this way.

Breast cancer is the second-leading cause of death among women worldwide. Early detection can be very effective, hence the importance of x-ray diagnostics. A mammogram cannot diagnose a malignant tumor, only give evidence of a lump or region of increased density within the breast. X-ray absorption by different types of soft tissue is very similar, so contrast is difficult; this is especially true for younger women, who typically have denser breasts. For older women who are at greater risk of developing breast cancer, the presence of more fat in the breast gives the lump or tumor more contrast. MRI (Magnetic resonance imaging) has recently been used as a supplement to conventional x rays to improve detection and eliminate false positives. The subject's radiation dose from x rays will be treated in a later chapter.

A standard x ray gives only a two-dimensional view of the object. Dense bones might hide images of soft tissue or organs. If you took another x ray from the side of the person (the first one being from the front), you would gain additional information. While shadow images are sufficient in many applications, far more sophisticated images can be produced with modern

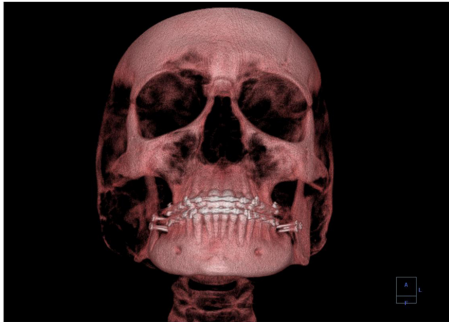
technology. [\[link\]](#) shows the use of a computed tomography (CT) scanner, also called computed axial tomography (CAT) scanner. X rays are passed through a narrow section (called a slice) of the patient's body (or body part) over a range of directions. An array of many detectors on the other side of the patient registers the x rays. The system is then rotated around the patient and another image is taken, and so on. The x-ray tube and detector array are mechanically attached and so rotate together. Complex computer image processing of the relative absorption of the x rays along different directions produces a highly-detailed image. Different slices are taken as the patient moves through the scanner on a table. Multiple images of different slices can also be computer analyzed to produce three-dimensional information, sometimes enhancing specific types of tissue, as shown in [\[link\]](#). G. Hounsfield (UK) and A. Cormack (US) won the Nobel Prize in Medicine in 1979 for their development of computed tomography.



A patient being positioned in a CT scanner aboard the hospital ship USNS Mercy. The CT scanner passes x rays through slices of the patient's body (or body part) over a range of directions.

The relative absorption of the x rays along different

directions is computer analyzed to produce highly detailed images. Three-dimensional information can be obtained from multiple slices. (credit: Rebecca Moat, U.S. Navy)



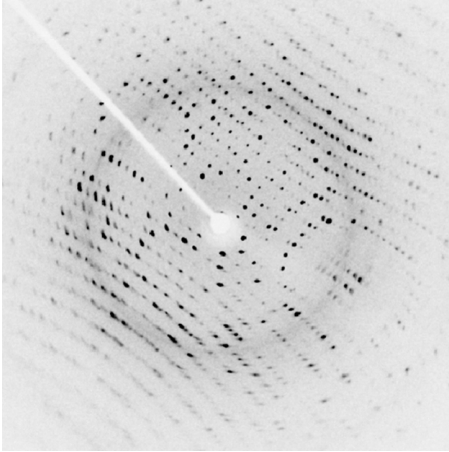
This three-dimensional image of a skull was produced by computed tomography, involving analysis of several x-ray slices of the head. (credit: Emailshankar, Wikimedia Commons)

X-Ray Diffraction and Crystallography

Since x-ray photons are very energetic, they have relatively short wavelengths. For example, the 54.4-keV K_{α} x ray of [\[link\]](#) has a

wavelength $\lambda = hc/E = 0.0228 \text{ nm}$. Thus, typical x-ray photons act like rays when they encounter macroscopic objects, like teeth, and produce sharp shadows; however, since atoms are on the order of 0.1 nm in size, x rays can be used to detect the location, shape, and size of atoms and molecules. The process is called **x-ray diffraction**, because it involves the diffraction and interference of x rays to produce patterns that can be analyzed for information about the structures that scattered the x rays. Perhaps the most famous example of x-ray diffraction is the discovery of the double-helix structure of DNA in 1953 by an international team of scientists working at the Cavendish Laboratory—American James Watson, Englishman Francis Crick, and New Zealand-born Maurice Wilkins. Using x-ray diffraction data produced by Rosalind Franklin, they were the first to discern the structure of DNA that is so crucial to life. For this, Watson, Crick, and Wilkins were awarded the 1962 Nobel Prize in Physiology or Medicine. There is much debate and controversy over the issue that Rosalind Franklin was not included in the prize.

[\[link\]](#) shows a diffraction pattern produced by the scattering of x rays from a crystal. This process is known as x-ray crystallography because of the information it can yield about crystal structure, and it was the type of data Rosalind Franklin supplied to Watson and Crick for DNA. Not only do x rays confirm the size and shape of atoms, they give information on the atomic arrangements in materials. For example, current research in high-temperature superconductors involves complex materials whose lattice arrangements are crucial to obtaining a superconducting material. These can be studied using x-ray crystallography.



X-ray diffraction from
the crystal of a protein,
hen egg lysozyme,
produced this
interference pattern.
Analysis of the pattern
yields information
about the structure of
the protein. (credit:
Del45, Wikimedia
Commons)

Historically, the scattering of x rays from crystals was used to prove that x rays are energetic EM waves. This was suspected from the time of the discovery of x rays in 1895, but it was not until 1912 that the German Max von Laue (1879–1960) convinced two of his colleagues to scatter x rays from crystals. If a diffraction pattern is obtained, he reasoned, then the x rays must be waves, and their wavelength could be determined. (The spacing of atoms in various crystals was reasonably well known at the time, based on good values for Avogadro's number.) The experiments were convincing, and the 1914 Nobel Prize in Physics was given to von Laue for his suggestion leading to the proof that x rays are EM waves. In 1915, the unique father-and-son team of Sir William Henry Bragg and his son Sir William Lawrence Bragg were awarded a joint Nobel Prize for inventing

the x-ray spectrometer and the then-new science of x-ray analysis. The elder Bragg had migrated to Australia from England just after graduating in mathematics. He learned physics and chemistry during his career at the University of Adelaide. The younger Bragg was born in Adelaide but went back to the Cavendish Laboratories in England to a career in x-ray and neutron crystallography; he provided support for Watson, Crick, and Wilkins for their work on unraveling the mysteries of DNA and to Max Perutz for his 1962 Nobel Prize-winning work on the structure of hemoglobin. Here again, we witness the enabling nature of physics—establishing instruments and designing experiments as well as solving mysteries in the biomedical sciences.

Certain other uses for x rays will be studied in later chapters. X rays are useful in the treatment of cancer because of the inhibiting effect they have on cell reproduction. X rays observed coming from outer space are useful in determining the nature of their sources, such as neutron stars and possibly black holes. Created in nuclear bomb explosions, x rays can also be used to detect clandestine atmospheric tests of these weapons. X rays can cause excitations of atoms, which then fluoresce (emitting characteristic EM radiation), making x-ray-induced fluorescence a valuable analytical tool in a range of fields from art to archaeology.

Section Summary

- X rays are relatively high-frequency EM radiation. They are produced by transitions between inner-shell electron levels, which produce x rays characteristic of the atomic element, or by decelerating electrons.
- X rays have many uses, including medical diagnostics and x-ray diffraction.

Conceptual Questions

Exercise:

Problem:

Explain why characteristic x rays are the most energetic in the EM emission spectrum of a given element.

Exercise:**Problem:**

Why does the energy of characteristic x rays become increasingly greater for heavier atoms?

Exercise:**Problem:**

Observers at a safe distance from an atmospheric test of a nuclear bomb feel its heat but receive none of its copious x rays. Why is air opaque to x rays but transparent to infrared?

Exercise:**Problem:**

Lasers are used to burn and read CDs. Explain why a laser that emits blue light would be capable of burning and reading more information than one that emits infrared.

Exercise:**Problem:**

Crystal lattices can be examined with x rays but not UV. Why?

Exercise:**Problem:**

CT scanners do not detect details smaller than about 0.5 mm. Is this limitation due to the wavelength of x rays? Explain.

Problem Exercises

Exercise:**Problem:**

(a) What is the shortest-wavelength x-ray radiation that can be generated in an x-ray tube with an applied voltage of 50.0 kV? (b) Calculate the photon energy in eV. (c) Explain the relationship of the photon energy to the applied voltage.

Solution:

(a) $0.248 \times 10^{-10} \text{ m}$

(b) 50.0 keV

(c) The photon energy is simply the applied voltage times the electron charge, so the value of the voltage in volts is the same as the value of the energy in electron volts.

Exercise:**Problem:**

A color television tube also generates some x rays when its electron beam strikes the screen. What is the shortest wavelength of these x rays, if a 30.0-kV potential is used to accelerate the electrons? (Note that TVs have shielding to prevent these x rays from exposing viewers.)

Exercise:**Problem:**

An x ray tube has an applied voltage of 100 kV. (a) What is the most energetic x-ray photon it can produce? Express your answer in electron volts and joules. (b) Find the wavelength of such an X-ray.

Solution:

(a) $100 \times 10^3 \text{ eV}$, $1.60 \times 10^{-14} \text{ J}$

(b) $0.124 \times 10^{-10} \text{ m}$

Exercise:

Problem:

The maximum characteristic x-ray photon energy comes from the capture of a free electron into a K shell vacancy. What is this photon energy in keV for tungsten, assuming the free electron has no initial kinetic energy?

Exercise:

Problem:

What are the approximate energies of the K_α and K_β x rays for copper?

Solution:

(a) 8.00 keV

(b) 9.48 keV

Glossary

x rays

a form of electromagnetic radiation

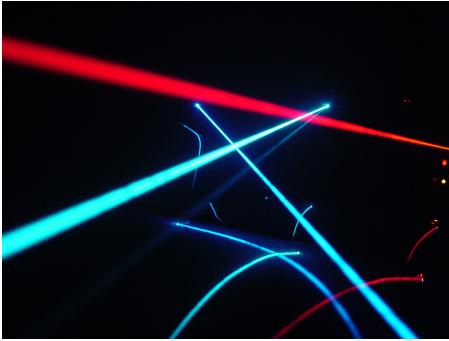
x-ray diffraction

a technique that provides the detailed information about crystallographic structure of natural and manufactured materials

Applications of Atomic Excitations and De-Excitations

- Define and discuss fluorescence.
- Define metastable.
- Describe how laser emission is produced.
- Explain population inversion.
- Define and discuss holography.

Many properties of matter and phenomena in nature are directly related to atomic energy levels and their associated excitations and de-excitations. The color of a rose, the output of a laser, and the transparency of air are but a few examples. (See [\[link\]](#).) While it may not appear that glow-in-the-dark pajamas and lasers have much in common, they are in fact different applications of similar atomic de-excitations.



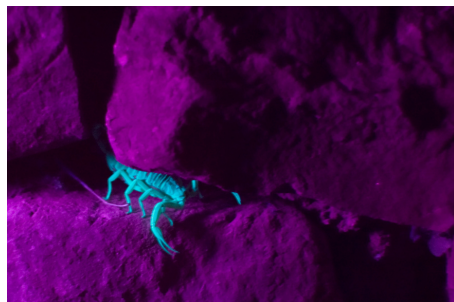
Light from a laser is based on a particular type of atomic de-excitation. (credit: Jeff Keyzer)

The color of a material is due to the ability of its atoms to absorb certain wavelengths while reflecting or reemitting others. A simple red material, for example a tomato, absorbs all visible wavelengths except red. This is because the atoms of its hydrocarbon pigment (lycopene) have levels separated by a variety of energies corresponding to all visible photon energies except red. Air is another interesting example. It is transparent to visible light, because there are few energy levels that visible photons can

excite in air molecules and atoms. Visible light, thus, cannot be absorbed. Furthermore, visible light is only weakly scattered by air, because visible wavelengths are so much greater than the sizes of the air molecules and atoms. Light must pass through kilometers of air to scatter enough to cause red sunsets and blue skies.

Fluorescence and Phosphorescence

The ability of a material to emit various wavelengths of light is similarly related to its atomic energy levels. [\[link\]](#) shows a scorpion illuminated by a UV lamp, sometimes called a black light. Some rocks also glow in black light, the particular colors being a function of the rock's mineral composition. Black lights are also used to make certain posters glow.



Objects glow in the visible spectrum when illuminated by an ultraviolet (black) light. Emissions are characteristic of the mineral involved, since they are related to its energy levels. In the case of scorpions, proteins near the surface of their skin give off the characteristic blue

glow. This is a colorful example of fluorescence in which excitation is induced by UV radiation while de-excitation occurs in the form of visible light. (credit: Ken Bosma, Flickr)

In the fluorescence process, an atom is excited to a level several steps above its ground state by the absorption of a relatively high-energy UV photon. This is called **atomic excitation**. Once it is excited, the atom can de-excite in several ways, one of which is to re-emit a photon of the same energy as excited it, a single step back to the ground state. This is called **atomic de-excitation**. All other paths of de-excitation involve smaller steps, in which lower-energy (longer wavelength) photons are emitted. Some of these may be in the visible range, such as for the scorpion in [\[link\]](#). **Fluorescence** is defined to be any process in which an atom or molecule, excited by a photon of a given energy, and de-excites by emission of a lower-energy photon.

Fluorescence can be induced by many types of energy input. Fluorescent paint, dyes, and even soap residues in clothes make colors seem brighter in sunlight by converting some UV into visible light. X rays can induce fluorescence, as is done in x-ray fluoroscopy to make brighter visible images. Electric discharges can induce fluorescence, as in so-called neon lights and in gas-discharge tubes that produce atomic and molecular spectra. Common fluorescent lights use an electric discharge in mercury vapor to cause atomic emissions from mercury atoms. The inside of a fluorescent light is coated with a fluorescent material that emits visible light over a broad spectrum of wavelengths. By choosing an appropriate coating, fluorescent lights can be made more like sunlight or like the reddish glow of candlelight, depending on needs. Fluorescent lights are more efficient in converting electrical energy into visible light than incandescent filaments

(about four times as efficient), the blackbody radiation of which is primarily in the infrared due to temperature limitations.

This atom is excited to one of its higher levels by absorbing a UV photon. It can de-excite in a single step, re-emitting a photon of the same energy, or in several steps. The process is called fluorescence if the atom de-excites in smaller steps, emitting energy different from that which excited it. Fluorescence can be induced by a variety of energy inputs, such as UV, x-rays, and electrical discharge.

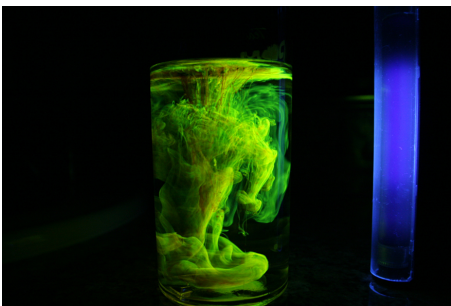
The spectacular Waitomo caves on North Island in New Zealand provide a natural habitat for glow-worms. The glow-worms hang up to 70 silk threads of about 30 or 40 cm each to trap prey that fly towards them in the dark. The fluorescence process is very efficient, with nearly 100% of the energy input turning into light. (In comparison, fluorescent lights are about 20% efficient.)

Fluorescence has many uses in biology and medicine. It is commonly used to label and follow a molecule within a cell. Such tagging allows one to study the structure of DNA and proteins. Fluorescent dyes and antibodies are usually used to tag the molecules, which are then illuminated with UV light and their emission of visible light is observed. Since the fluorescence of each element is characteristic, identification of elements within a sample can be done this way.

[\[link\]](#) shows a commonly used fluorescent dye called fluorescein. Below that, [\[link\]](#) reveals the diffusion of a fluorescent dye in water by observing it under UV light.



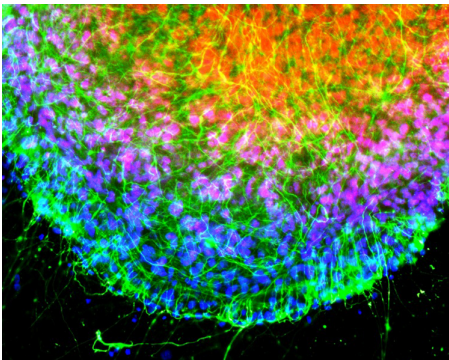
Fluorescein, shown here in powder form, is used to dye laboratory samples.
(credit: Benjah-bmm27, Wikimedia Commons)



Here, fluorescent powder is added to a beaker of water. The mixture gives off a bright glow under ultraviolet light.
(credit: Bricksnite, Wikimedia Commons)

Note:**Nano-Crystals**

Recently, a new class of fluorescent materials has appeared—“nano-crystals.” These are single-crystal molecules less than 100 nm in size. The smallest of these are called “quantum dots.” These semiconductor indicators are very small (2–6 nm) and provide improved brightness. They also have the advantage that all colors can be excited with the same incident wavelength. They are brighter and more stable than organic dyes and have a longer lifetime than conventional phosphors. They have become an excellent tool for long-term studies of cells, including migration and morphology. ([link](#).)



Microscopic image of chicken cells using nano-crystals of a fluorescent dye. Cell nuclei exhibit blue fluorescence while neurofilaments exhibit green. (credit: Weerapong Prasongchean, Wikimedia Commons)

Once excited, an atom or molecule will usually spontaneously de-excite quickly. (The electrons raised to higher levels are attracted to lower ones by the positive charge of the nucleus.) Spontaneous de-excitation has a very short mean lifetime of typically about 10^{-8} s. However, some levels have significantly longer lifetimes, ranging up to milliseconds to minutes or even hours. These energy levels are inhibited and are slow in de-exciting because their quantum numbers differ greatly from those of available lower levels. Although these level lifetimes are short in human terms, they are many orders of magnitude longer than is typical and, thus, are said to be **metastable**, meaning relatively stable. **Phosphorescence** is the de-excitation of a metastable state. Glow-in-the-dark materials, such as luminous dials on some watches and clocks and on children's toys and pajamas, are made of phosphorescent substances. Visible light excites the atoms or molecules to metastable states that decay slowly, releasing the stored excitation energy partially as visible light. In some ceramics, atomic excitation energy can be frozen in after the ceramic has cooled from its firing. It is very slowly released, but the ceramic can be induced to phosphoresce by heating—a process called “thermoluminescence.” Since the release is slow, thermoluminescence can be used to date antiquities. The less light emitted, the older the ceramic. (See [\[link\]](#).)



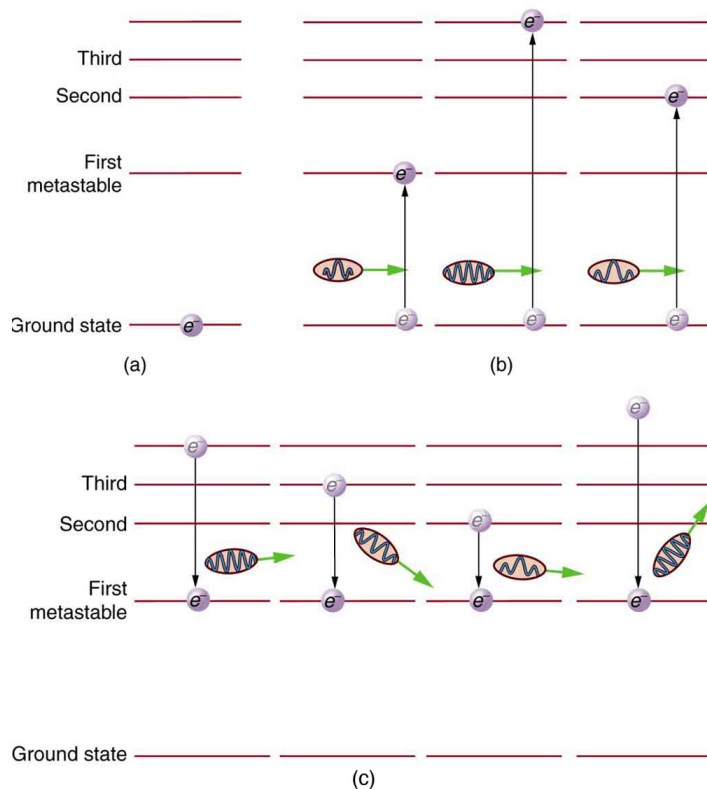
Atoms frozen in an excited state when this Chinese ceramic figure was fired can be stimulated to de-excite and emit EM radiation by heating a sample of the ceramic—a process called thermoluminescence. Since the states slowly de-excite over centuries, the amount of thermoluminescence decreases with age, making it possible to use this effect to date and authenticate antiquities. This figure dates from the 11th century. (credit: Vassil, Wikimedia Commons)

Lasers

Lasers today are commonplace. Lasers are used to read bar codes at stores and in libraries, laser shows are staged for entertainment, laser printers produce high-quality images at relatively low cost, and lasers send prodigious numbers of telephone messages through optical fibers. Among other things, lasers are also employed in surveying, weapons guidance,

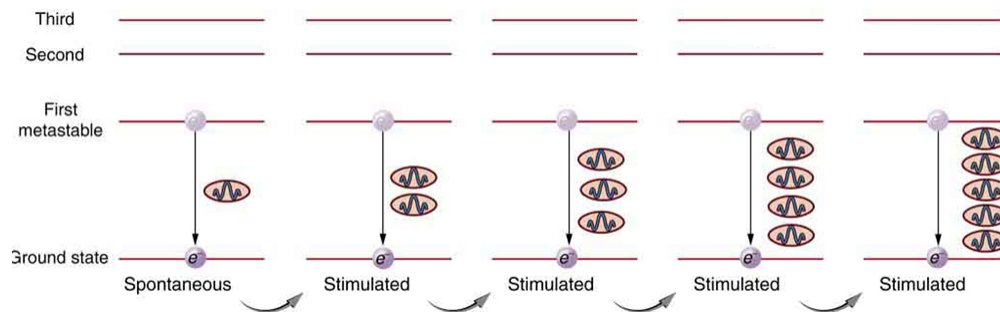
tumor eradication, retinal welding, and for reading music CDs and computer CD-ROMs.

Why do lasers have so many varied applications? The answer is that lasers produce single-wavelength EM radiation that is also very coherent—that is, the emitted photons are in phase. Laser output can, thus, be more precisely manipulated than incoherent mixed-wavelength EM radiation from other sources. The reason laser output is so pure and coherent is based on how it is produced, which in turn depends on a metastable state in the lasing material. Suppose a material had the energy levels shown in [\[link\]](#). When energy is put into a large collection of these atoms, electrons are raised to all possible levels. Most return to the ground state in less than about 10^{-8} s, but those in the metastable state linger. This includes those electrons originally excited to the metastable state and those that fell into it from above. It is possible to get a majority of the atoms into the metastable state, a condition called a **population inversion**.



(a) Energy-level diagram for an atom showing the first few states, one of which is metastable. (b) Massive energy input excites atoms to a variety of states. (c) Most states decay quickly, leaving electrons only in the metastable and ground state. If a majority of electrons are in the metastable state, a population inversion has been achieved.

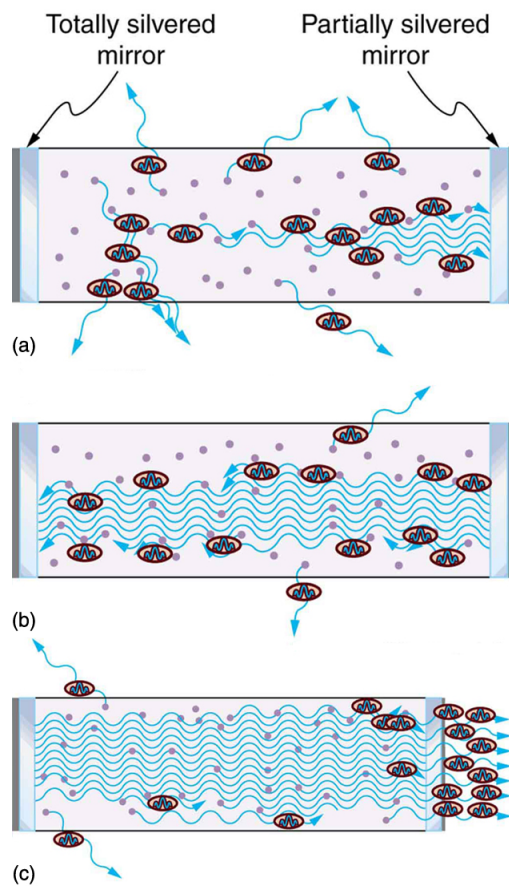
Once a population inversion is achieved, a very interesting thing can happen, as shown in [\[link\]](#). An electron spontaneously falls from the metastable state, emitting a photon. This photon finds another atom in the metastable state and stimulates it to decay, emitting a second photon of *the same wavelength and in phase* with the first, and so on. **Stimulated emission** is the emission of electromagnetic radiation in the form of photons of a given frequency, triggered by photons of the same frequency. For example, an excited atom, with an electron in an energy orbit higher than normal, releases a photon of a specific frequency when the electron drops back to a lower energy orbit. If this photon then strikes another electron in the same high-energy orbit in another atom, another photon of the same frequency is released. The emitted photons and the triggering photons are always in phase, have the same polarization, and travel in the same direction. The probability of absorption of a photon is the same as the probability of stimulated emission, and so a majority of atoms must be in the metastable state to produce energy. Einstein (again Einstein, and back in 1917!) was one of the important contributors to the understanding of stimulated emission of radiation. Among other things, Einstein was the first to realize that stimulated emission and absorption are equally probable. The laser acts as a temporary energy storage device that subsequently produces a massive energy output of single-wavelength, in-phase photons.



One atom in the metastable state spontaneously decays to a lower level, producing a photon that goes on to stimulate another atom to de-excite. The second photon has exactly the same energy and wavelength as the first and is in phase with it. Both go on to stimulate the emission of other photons. A population inversion is necessary for there to be a net production rather than a net absorption of the photons.

The name **laser** is an acronym for light amplification by stimulated emission of radiation, the process just described. The process was proposed and developed following the advances in quantum physics. A joint Nobel Prize was awarded in 1964 to American Charles Townes (1915–), and Nikolay Basov (1922–2001) and Aleksandr Prokhorov (1916–2002), from the Soviet Union, for the development of lasers. The Nobel Prize in 1981 went to Arthur Schawlow (1921-1999) for pioneering laser applications. The original devices were called masers, because they produced microwaves. The first working laser was created in 1960 at Hughes Research labs (CA) by T. Maiman. It used a pulsed high-powered flash lamp and a ruby rod to produce red light. Today the name laser is used for all such devices developed to produce a variety of wavelengths, including microwave, infrared, visible, and ultraviolet radiation. [\[link\]](#) shows how a laser can be constructed to enhance the stimulated emission of radiation. Energy input can be from a flash tube, electrical discharge, or other sources, in a process sometimes called optical pumping. A large percentage of the original pumping energy is dissipated in other forms, but a population inversion must be achieved. Mirrors can be used to enhance stimulated

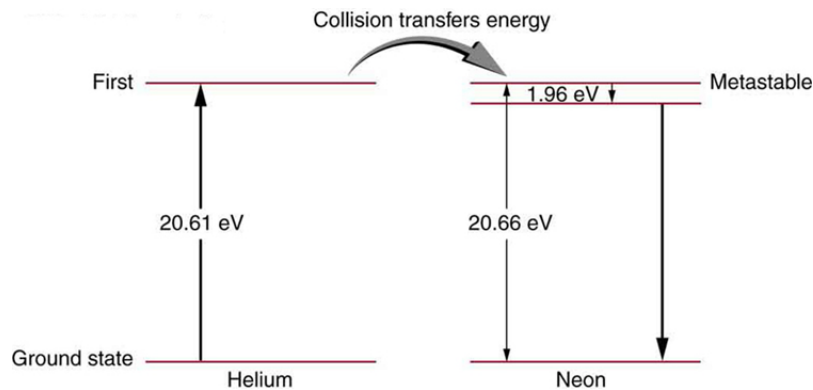
emission by multiple passes of the radiation back and forth through the lasing material. One of the mirrors is semitransparent to allow some of the light to pass through. The laser output from a laser is a mere 1% of the light passing back and forth in a laser.



Typical laser construction has a method of pumping energy into the lasing material to produce a population inversion. (a) Spontaneous emission begins with some photons escaping and others stimulating further emissions. (b) and (c)

Mirrors are used to enhance the probability of stimulated emission by passing photons through the material several times.

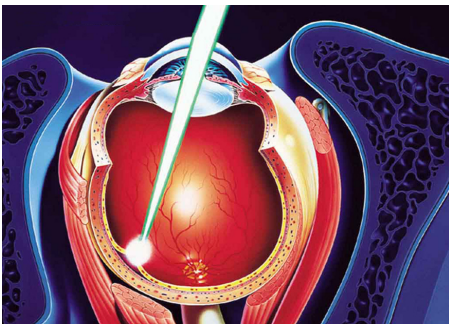
Lasers are constructed from many types of lasing materials, including gases, liquids, solids, and semiconductors. But all lasers are based on the existence of a metastable state or a phosphorescent material. Some lasers produce continuous output; others are pulsed in bursts as brief as 10^{-14} s. Some laser outputs are fantastically powerful—some greater than 10^{12} W—but the more common, everyday lasers produce something on the order of 10^{-3} W. The helium-neon laser that produces a familiar red light is very common. [\[link\]](#) shows the energy levels of helium and neon, a pair of noble gases that work well together. An electrical discharge is passed through a helium-neon gas mixture in which the number of atoms of helium is ten times that of neon. The first excited state of helium is metastable and, thus, stores energy. This energy is easily transferred by collision to neon atoms, because they have an excited state at nearly the same energy as that in helium. That state in neon is also metastable, and this is the one that produces the laser output. (The most likely transition is to the nearby state, producing 1.96 eV photons, which have a wavelength of 633 nm and appear red.) A population inversion can be produced in neon, because there are so many more helium atoms and these put energy into the neon. Helium-neon lasers often have continuous output, because the population inversion can be maintained even while lasing occurs. Probably the most common lasers in use today, including the common laser pointer, are semiconductor or diode lasers, made of silicon. Here, energy is pumped into the material by passing a current in the device to excite the electrons. Special coatings on the ends and fine cleavings of the semiconductor material allow light to bounce back and forth and a tiny fraction to emerge as laser light. Diode lasers can usually run continually and produce outputs in the milliwatt range.



Energy levels in helium and neon. In the common helium-neon laser, an electrical discharge pumps energy into the metastable states of both atoms. The gas mixture has about ten times more helium atoms than neon atoms. Excited helium atoms easily de-excite by transferring energy to neon in a collision. A population inversion in neon is achieved, allowing lasing by the neon to occur.

There are many medical applications of lasers. Lasers have the advantage that they can be focused to a small spot. They also have a well-defined wavelength. Many types of lasers are available today that provide wavelengths from the ultraviolet to the infrared. This is important, as one needs to be able to select a wavelength that will be preferentially absorbed by the material of interest. Objects appear a certain color because they absorb all other visible colors incident upon them. What wavelengths are absorbed depends upon the energy spacing between electron orbitals in that molecule. Unlike the hydrogen atom, biological molecules are complex and have a variety of absorption wavelengths or lines. But these can be determined and used in the selection of a laser with the appropriate wavelength. Water is transparent to the visible spectrum but will absorb light in the UV and IR regions. Blood (hemoglobin) strongly reflects red but absorbs most strongly in the UV.

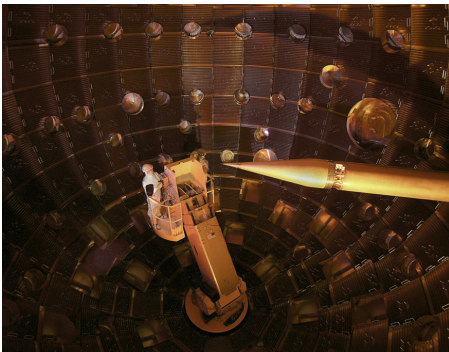
Laser surgery uses a wavelength that is strongly absorbed by the tissue it is focused upon. One example of a medical application of lasers is shown in [\[link\]](#). A detached retina can result in total loss of vision. Burns made by a laser focused to a small spot on the retina form scar tissue that can hold the retina in place, salvaging the patient's vision. Other light sources cannot be focused as precisely as a laser due to refractive dispersion of different wavelengths. Similarly, laser surgery in the form of cutting or burning away tissue is made more accurate because laser output can be very precisely focused and is preferentially absorbed because of its single wavelength. Depending upon what part or layer of the retina needs repairing, the appropriate type of laser can be selected. For the repair of tears in the retina, a green argon laser is generally used. This light is absorbed well by tissues containing blood, so coagulation or "welding" of the tear can be done.



A detached retina is burned by a laser designed to focus on a small spot on the retina, the resulting scar tissue holding it in place. The lens of the eye is used to focus the light, as is the device bringing the laser output to the eye.

In dentistry, the use of lasers is rising. Lasers are most commonly used for surgery on the soft tissue of the mouth. They can be used to remove ulcers, stop bleeding, and reshape gum tissue. Their use in cutting into bones and teeth is not quite so common; here the erbium YAG (yttrium aluminum garnet) laser is used.

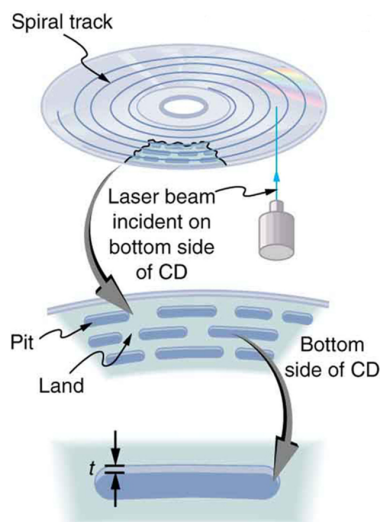
The massive combination of lasers shown in [\[link\]](#) can be used to induce nuclear fusion, the energy source of the sun and hydrogen bombs. Since lasers can produce very high power in very brief pulses, they can be used to focus an enormous amount of energy on a small glass sphere containing fusion fuel. Not only does the incident energy increase the fuel temperature significantly so that fusion can occur, it also compresses the fuel to great density, enhancing the probability of fusion. The compression or implosion is caused by the momentum of the impinging laser photons.



This system of lasers
at Lawrence
Livermore Laboratory
is used to ignite
nuclear fusion. A
tremendous burst of
energy is focused on a
small fuel pellet,
which is imploded to
the high density and
temperature needed to
make the fusion

reaction proceed.
(credit: Lawrence
Livermore National
Laboratory, Lawrence
Livermore National
Security, LLC, and the
Department of
Energy)

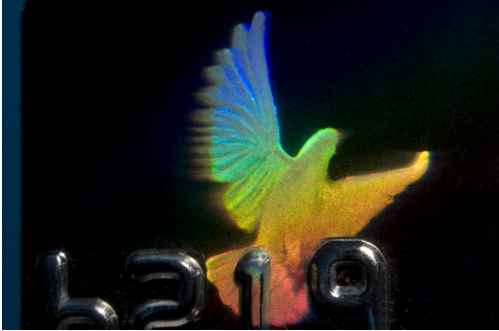
Music CDs are now so common that vinyl records are quaint antiquities. CDs (and DVDs) store information digitally and have a much larger information-storage capacity than vinyl records. An entire encyclopedia can be stored on a single CD. [\[link\]](#) illustrates how the information is stored and read from the CD. Pits made in the CD by a laser can be tiny and very accurately spaced to record digital information. These are read by having an inexpensive solid-state infrared laser beam scatter from pits as the CD spins, revealing their digital pattern and the information encoded upon them.



A CD has digital
information

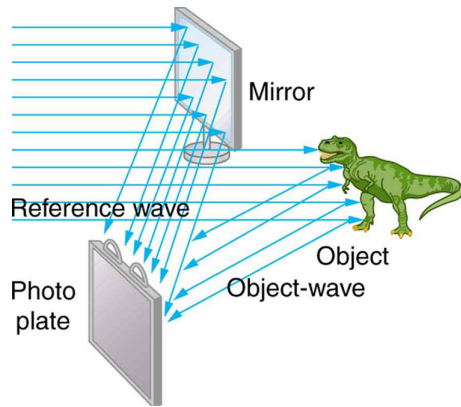
stored in the form
of laser-created
pits on its
surface. These in
turn can be read
by detecting the
laser light
scattered from the
pit. Large
information
capacity is
possible because
of the precision
of the laser.
Shorter-
wavelength lasers
enable greater
storage capacity.

Holograms, such as those in [\[link\]](#), are true three-dimensional images recorded on film by lasers. Holograms are used for amusement, decoration on novelty items and magazine covers, security on credit cards and driver's licenses (a laser and other equipment is needed to reproduce them), and for serious three-dimensional information storage. You can see that a hologram is a true three-dimensional image, because objects change relative position in the image when viewed from different angles.



Credit cards commonly have holograms for logos, making them difficult to reproduce (credit: Dominic Alves, Flickr)

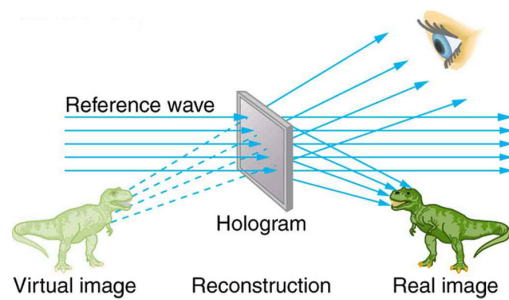
The name **hologram** means “entire picture” (from the Greek *holo*, as in holistic), because the image is three-dimensional. **Holography** is the process of producing holograms and, although they are recorded on photographic film, the process is quite different from normal photography. Holography uses light interference or wave optics, whereas normal photography uses geometric optics. [\[link\]](#) shows one method of producing a hologram. Coherent light from a laser is split by a mirror, with part of the light illuminating the object. The remainder, called the reference beam, shines directly on a piece of film. Light scattered from the object interferes with the reference beam, producing constructive and destructive interference. As a result, the exposed film looks foggy, but close examination reveals a complicated interference pattern stored on it. Where the interference was constructive, the film (a negative actually) is darkened. Holography is sometimes called lensless photography, because it uses the wave characteristics of light as contrasted to normal photography, which uses geometric optics and so requires lenses.



Production of a hologram. Single-wavelength coherent light from a laser produces a well-defined interference pattern on a piece of film. The laser beam is split by a partially silvered mirror, with part of the light illuminating the object and the remainder shining directly on the film.

Light falling on a hologram can form a three-dimensional image. The process is complicated in detail, but the basics can be understood as shown in [\[link\]](#), in which a laser of the same type that exposed the film is now used to illuminate it. The myriad tiny exposed regions of the film are dark and block the light, while less exposed regions allow light to pass. The film thus acts much like a collection of diffraction gratings with various spacings. Light passing through the hologram is diffracted in various directions, producing both real and virtual images of the object used to expose the film. The interference pattern is the same as that produced by the object. Moving

your eye to various places in the interference pattern gives you different perspectives, just as looking directly at the object would. The image thus looks like the object and is three-dimensional like the object.



A transmission hologram is one that produces real and virtual images when a laser of the same type as that which exposed the hologram is passed through it. Diffraction from various parts of the film produces the same interference pattern as the object that was used to expose it.

The hologram illustrated in [\[link\]](#) is a transmission hologram. Holograms that are viewed with reflected light, such as the white light holograms on credit cards, are reflection holograms and are more common. White light holograms often appear a little blurry with rainbow edges, because the diffraction patterns of various colors of light are at slightly different locations due to their different wavelengths. Further uses of holography include all types of 3-D information storage, such as of statues in museums and engineering studies of structures and 3-D images of human organs. Invented in the late 1940s by Dennis Gabor (1900–1970), who won the

1971 Nobel Prize in Physics for his work, holography became far more practical with the development of the laser. Since lasers produce coherent single-wavelength light, their interference patterns are more pronounced. The precision is so great that it is even possible to record numerous holograms on a single piece of film by just changing the angle of the film for each successive image. This is how the holograms that move as you walk by them are produced—a kind of lensless movie.

In a similar way, in the medical field, holograms have allowed complete 3-D holographic displays of objects from a stack of images. Storing these images for future use is relatively easy. With the use of an endoscope, high-resolution 3-D holographic images of internal organs and tissues can be made.

Section Summary

- An important atomic process is fluorescence, defined to be any process in which an atom or molecule is excited by absorbing a photon of a given energy and de-excited by emitting a photon of a lower energy.
- Some states live much longer than others and are termed metastable.
- Phosphorescence is the de-excitation of a metastable state.
- Lasers produce coherent single-wavelength EM radiation by stimulated emission, in which a metastable state is stimulated to decay.
- Lasing requires a population inversion, in which a majority of the atoms or molecules are in their metastable state.

Conceptual Questions

Exercise:

Problem:

How do the allowed orbits for electrons in atoms differ from the allowed orbits for planets around the sun? Explain how the correspondence principle applies here.

Exercise:

Problem:

Atomic and molecular spectra are discrete. What does discrete mean, and how are discrete spectra related to the quantization of energy and electron orbits in atoms and molecules?

Exercise:**Problem:**

Hydrogen gas can only absorb EM radiation that has an energy corresponding to a transition in the atom, just as it can only emit these discrete energies. When a spectrum is taken of the solar corona, in which a broad range of EM wavelengths are passed through very hot hydrogen gas, the absorption spectrum shows all the features of the emission spectrum. But when such EM radiation passes through room-temperature hydrogen gas, only the Lyman series is absorbed. Explain the difference.

Exercise:**Problem:**

Lasers are used to burn and read CDs. Explain why a laser that emits blue light would be capable of burning and reading more information than one that emits infrared.

Exercise:**Problem:**

The coating on the inside of fluorescent light tubes absorbs ultraviolet light and subsequently emits visible light. An inventor claims that he is able to do the reverse process. Is the inventor's claim possible?

Exercise:**Problem:**

What is the difference between fluorescence and phosphorescence?

Exercise:

Problem:

How can you tell that a hologram is a true three-dimensional image and that those in 3-D movies are not?

Problem Exercises**Exercise:****Problem:**

[\[link\]](#) shows the energy-level diagram for neon. (a) Verify that the energy of the photon emitted when neon goes from its metastable state to the one immediately below is equal to 1.96 eV. (b) Show that the wavelength of this radiation is 633 nm. (c) What wavelength is emitted when the neon makes a direct transition to its ground state?

Solution:

(a) 1.96 eV

(b) $(1240 \text{ eV}\cdot\text{nm})/(1.96 \text{ eV}) = 633 \text{ nm}$

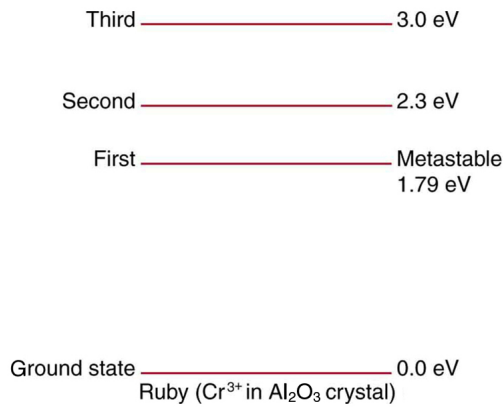
(c) 60.0 nm

Exercise:**Problem:**

A helium-neon laser is pumped by electric discharge. What wavelength electromagnetic radiation would be needed to pump it? See [\[link\]](#) for energy-level information.

Exercise:**Problem:**

Ruby lasers have chromium atoms doped in an aluminum oxide crystal. The energy level diagram for chromium in a ruby is shown in [\[link\]](#). What wavelength is emitted by a ruby laser?



Chromium atoms in an aluminum oxide crystal have these energy levels, one of which is metastable. This is the basis of a ruby laser. Visible light can pump the atom into an excited state above the metastable state to achieve a population inversion.

Solution:

693 nm

Exercise:

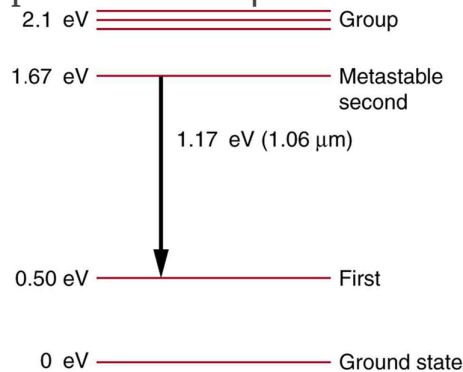
Problem:

(a) What energy photons can pump chromium atoms in a ruby laser from the ground state to its second and third excited states? (b) What are the wavelengths of these photons? Verify that they are in the visible part of the spectrum.

Exercise:

Problem:

Some of the most powerful lasers are based on the energy levels of neodymium in solids, such as glass, as shown in [\[link\]](#). (a) What average wavelength light can pump the neodymium into the levels above its metastable state? (b) Verify that the 1.17 eV transition produces 1.06 μm radiation.



Neodymium atoms in glass have these energy levels, one of which is metastable.

The group of levels above the metastable state is convenient for achieving a population inversion, since photons of many different energies can be absorbed by atoms in the ground state.

Solution:

(a) 590 nm

(b) $(1240 \text{ eV}\cdot\text{nm})/(1.17 \text{ eV}) = 1.06 \mu\text{m}$

Glossary

metastable

a state whose lifetime is an order of magnitude longer than the most short-lived states

atomic excitation

a state in which an atom or ion acquires the necessary energy to promote one or more of its electrons to electronic states higher in energy than their ground state

atomic de-excitation

process by which an atom transfers from an excited electronic state back to the ground state electronic configuration; often occurs by emission of a photon

laser

acronym for light amplification by stimulated emission of radiation

phosphorescence

the de-excitation of a metastable state

population inversion

the condition in which the majority of atoms in a sample are in a metastable state

stimulated emission

emission by atom or molecule in which an excited state is stimulated to decay, most readily caused by a photon of the same energy that is necessary to excite the state

hologram

means *entire picture* (from the Greek word *holo*, as in holistic), because the image produced is three dimensional

holography

the process of producing holograms

fluorescence

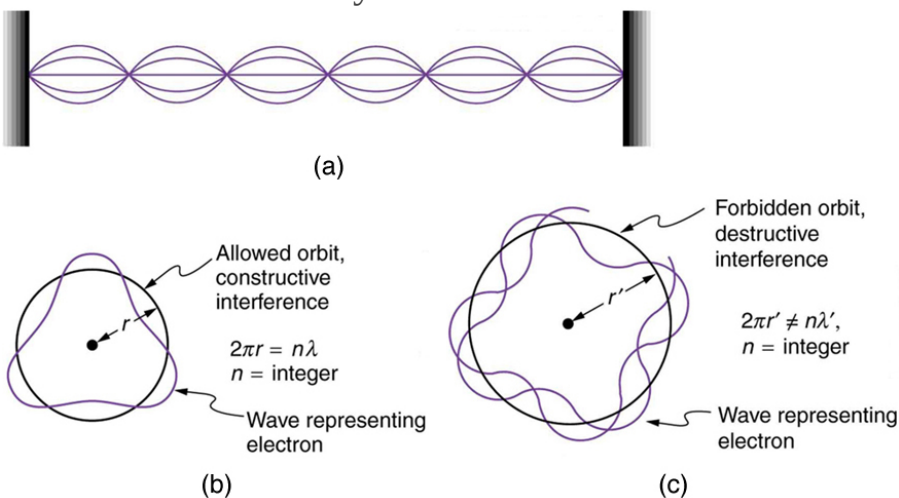
any process in which an atom or molecule, excited by a photon of a given energy, de-excites by emission of a lower-energy photon

The Wave Nature of Matter Causes Quantization

- Explain Bohr's model of atom.
- Define and describe quantization of angular momentum.
- Calculate the angular momentum for an orbit of atom.
- Define and describe the wave-like properties of matter.

After visiting some of the applications of different aspects of atomic physics, we now return to the basic theory that was built upon Bohr's atom. Einstein once said it was important to keep asking the questions we eventually teach children not to ask. Why is angular momentum quantized? You already know the answer. Electrons have wave-like properties, as de Broglie later proposed. They can exist only where they interfere constructively, and only certain orbits meet proper conditions, as we shall see in the next module.

Following Bohr's initial work on the hydrogen atom, a decade was to pass before de Broglie proposed that matter has wave properties. The wave-like properties of matter were subsequently confirmed by observations of electron interference when scattered from crystals. Electrons can exist only in locations where they interfere constructively. How does this affect electrons in atomic orbits? When an electron is bound to an atom, its wavelength must fit into a small space, something like a standing wave on a string. (See [\[link\]](#).) Allowed orbits are those orbits in which an electron constructively interferes with itself. Not all orbits produce constructive interference. Thus only certain orbits are allowed—the orbits are quantized.



(a) Waves on a string have a wavelength related to the length of the string, allowing them to interfere constructively. (b) If we imagine the string bent into a closed circle, we get a rough idea of how electrons in circular orbits can interfere constructively. (c) If the wavelength does not fit into the circumference, the electron interferes destructively; it cannot exist in such an orbit.

For a circular orbit, constructive interference occurs when the electron's wavelength fits neatly into the circumference, so that wave crests always align with crests and wave troughs align with troughs, as shown in [\[link\]](#) (b). More precisely, when an integral multiple of the electron's wavelength equals the circumference of the orbit, constructive interference is obtained. In equation form, the *condition for constructive interference and an allowed electron orbit* is

Equation:

$$n\lambda_n = 2\pi r_n (n = 1, 2, 3 \dots),$$

where λ_n is the electron's wavelength and r_n is the radius of that circular orbit. The de Broglie wavelength is $\lambda = h/p = h/mv$, and so here $\lambda = h/m_e v$. Substituting this into the previous condition for constructive interference produces an interesting result:

Equation:

$$\frac{nh}{m_e v} = 2\pi r_n.$$

Rearranging terms, and noting that $L = mvr$ for a circular orbit, we obtain the quantization of angular momentum as the condition for allowed orbits:

Equation:

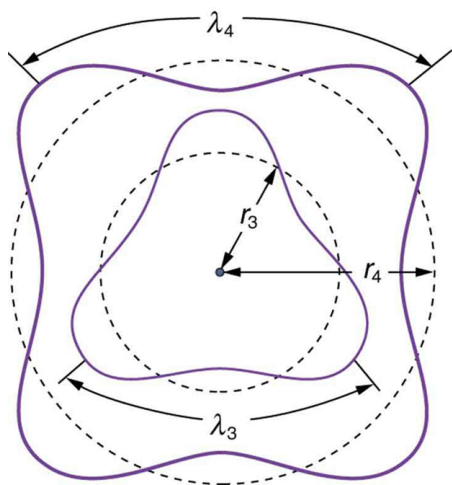
$$L = m_e v r_n = n \frac{h}{2\pi} (n = 1, 2, 3 \dots).$$

This is what Bohr was forced to hypothesize as the rule for allowed orbits, as stated earlier. We now realize that it is the condition for constructive interference of an electron in a circular orbit. [\[link\]](#) illustrates this for $n = 3$ and $n = 4$.

Note:

Waves and Quantization

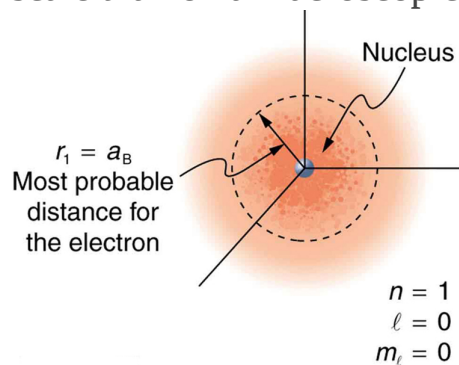
The wave nature of matter is responsible for the quantization of energy levels in bound systems. Only those states where matter interferes constructively exist, or are “allowed.” Since there is a lowest orbit where this is possible in an atom, the electron cannot spiral into the nucleus. It cannot exist closer to or inside the nucleus. The wave nature of matter is what prevents matter from collapsing and gives atoms their sizes.



The third and fourth
allowed circular orbits
have three and four
wavelengths,

respectively, in their
circumferences.

Because of the wave character of matter, the idea of well-defined orbits gives way to a model in which there is a cloud of probability, consistent with Heisenberg's uncertainty principle. [\[link\]](#) shows how this applies to the ground state of hydrogen. If you try to follow the electron in some well-defined orbit using a probe that has a small enough wavelength to get some details, you will instead knock the electron out of its orbit. Each measurement of the electron's position will find it to be in a definite location somewhere near the nucleus. Repeated measurements reveal a cloud of probability like that in the figure, with each speck the location determined by a single measurement. There is not a well-defined, circular-orbit type of distribution. Nature again proves to be different on a small scale than on a macroscopic scale.



The ground state of a
hydrogen atom has a
probability cloud
describing the position
of its electron. The
probability of finding
the electron is
proportional to the
darkness of the cloud.
The electron can be
closer or farther than

the Bohr radius, but it is very unlikely to be a great distance from the nucleus.

There are many examples in which the wave nature of matter causes quantization in bound systems such as the atom. Whenever a particle is confined or bound to a small space, its allowed wavelengths are those which fit into that space. For example, the particle in a box model describes a particle free to move in a small space surrounded by impenetrable barriers. This is true in blackbody radiators (atoms and molecules) as well as in atomic and molecular spectra. Various atoms and molecules will have different sets of electron orbits, depending on the size and complexity of the system. When a system is large, such as a grain of sand, the tiny particle waves in it can fit in so many ways that it becomes impossible to see that the allowed states are discrete. Thus the correspondence principle is satisfied. As systems become large, they gradually look less grainy, and quantization becomes less evident. Unbound systems (small or not), such as an electron freed from an atom, do not have quantized energies, since their wavelengths are not constrained to fit in a certain volume.

Note:

PhET Explorations: Quantum Wave Interference

When do photons, electrons, and atoms behave like particles and when do they behave like waves? Watch waves spread out and interfere as they pass through a double slit, then get detected on a screen as tiny dots. Use quantum detectors to explore how measurements change the waves and the patterns they produce on the screen.

[Quantum](#)
[Wave](#)

Section Summary

- Quantization of orbital energy is caused by the wave nature of matter. Allowed orbits in atoms occur for constructive interference of electrons in the orbit, requiring an integral number of wavelengths to fit in an orbit's circumference; that is,

Equation:

$$n\lambda_n = 2\pi r_n (n = 1, 2, 3 \dots),$$

where λ_n is the electron's de Broglie wavelength.

- Owing to the wave nature of electrons and the Heisenberg uncertainty principle, there are no well-defined orbits; rather, there are clouds of probability.
- Bohr correctly proposed that the energy and radii of the orbits of electrons in atoms are quantized, with energy for transitions between orbits given by

Equation:

$$\Delta E = hf = E_i - E_f,$$

where ΔE is the change in energy between the initial and final orbits and hf is the energy of an absorbed or emitted photon.

- It is useful to plot orbit energies on a vertical graph called an energy-level diagram.
- The allowed orbits are circular, Bohr proposed, and must have quantized orbital angular momentum given by

Equation:

$$L = m_e v r_n = n \frac{h}{2\pi} (n = 1, 2, 3 \dots),$$

where L is the angular momentum, r_n is the radius of orbit n , and h is Planck's constant.

Conceptual Questions

Exercise:

Problem:

How is the de Broglie wavelength of electrons related to the quantization of their orbits in atoms and molecules?

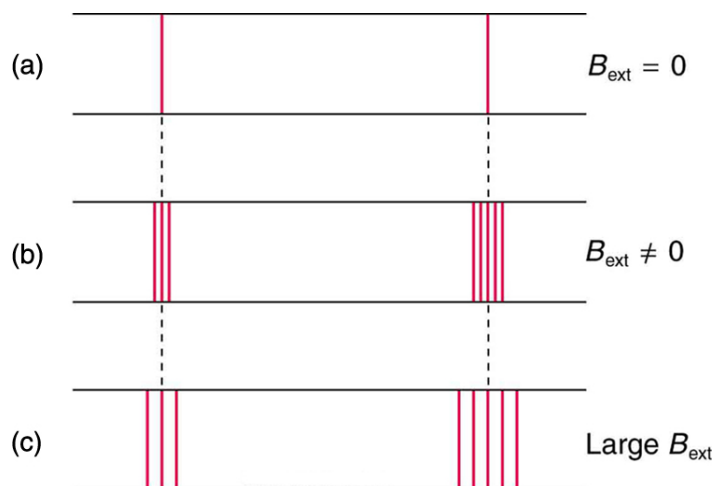
Patterns in Spectra Reveal More Quantization

- State and discuss the Zeeman effect.
- Define orbital magnetic field.
- Define orbital angular momentum.
- Define space quantization.

High-resolution measurements of atomic and molecular spectra show that the spectral lines are even more complex than they first appear. In this section, we will see that this complexity has yielded important new information about electrons and their orbits in atoms.

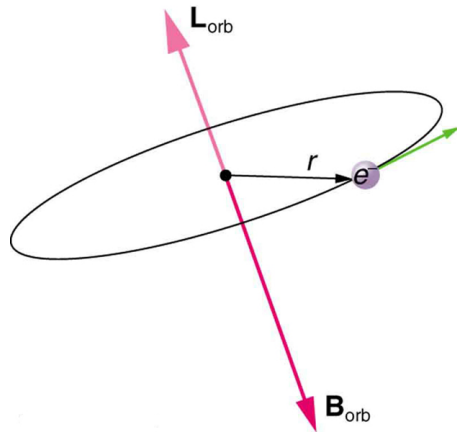
In order to explore the substructure of atoms (and knowing that magnetic fields affect moving charges), the Dutch physicist Hendrik Lorentz (1853–1930) suggested that his student Pieter Zeeman (1865–1943) study how spectra might be affected by magnetic fields. What they found became known as the **Zeeman effect**, which involved spectral lines being split into two or more separate emission lines by an external magnetic field, as shown in [\[link\]](#). For their discoveries, Zeeman and Lorentz shared the 1902 Nobel Prize in Physics.

Zeeman splitting is complex. Some lines split into three lines, some into five, and so on. But one general feature is that the amount the split lines are separated is proportional to the applied field strength, indicating an interaction with a moving charge. The splitting means that the quantized energy of an orbit is affected by an external magnetic field, causing the orbit to have several discrete energies instead of one. Even without an external magnetic field, very precise measurements showed that spectral lines are doublets (split into two), apparently by magnetic fields within the atom itself.

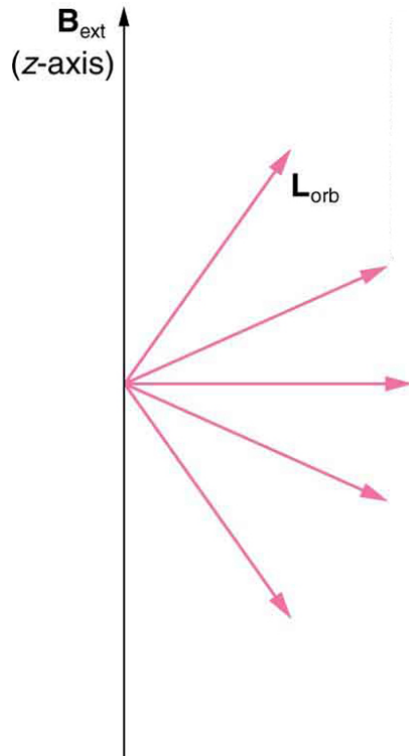


The Zeeman effect is the splitting of spectral lines when a magnetic field is applied. The number of lines formed varies, but the spread is proportional to the strength of the applied field. (a) Two spectral lines with no external magnetic field. (b) The lines split when the field is applied. (c) The splitting is greater when a stronger field is applied.

Bohr's theory of circular orbits is useful for visualizing how an electron's orbit is affected by a magnetic field. The circular orbit forms a current loop, which creates a magnetic field of its own, \mathbf{B}_{orb} as seen in [\[link\]](#). Note that the **orbital magnetic field** \mathbf{B}_{orb} and the **orbital angular momentum** \mathbf{L}_{orb} are along the same line. The external magnetic field and the orbital magnetic field interact; a torque is exerted to align them. A torque rotating a system through some angle does work so that there is energy associated with this interaction. Thus, orbits at different angles to the external magnetic field have different energies. What is remarkable is that the energies are quantized—the magnetic field splits the spectral lines into several discrete lines that have different energies. This means that only certain angles are allowed between the orbital angular momentum and the external field, as seen in [\[link\]](#).



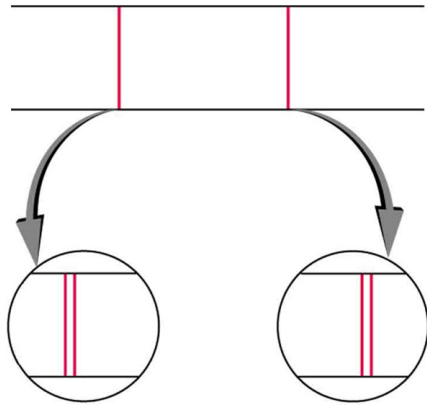
The approximate picture of an electron in a circular orbit illustrates how the current loop produces its own magnetic field, called \mathbf{B}_{orb} . It also shows how \mathbf{B}_{orb} is along the same line as the orbital angular momentum \mathbf{L}_{orb} .



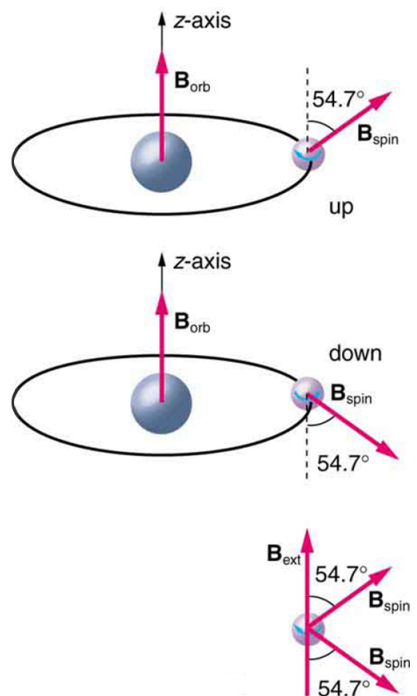
Only certain angles are allowed between the orbital angular momentum and an external magnetic field. This is implied by the fact that the Zeeman effect splits spectral lines into several discrete lines. Each line is associated with an angle between the external magnetic field and magnetic fields due to electrons and their orbits.

We already know that the magnitude of angular momentum is quantized for electron orbits in atoms. The new insight is that the *direction of the orbital angular momentum is also quantized*. The fact that the orbital angular momentum can have only certain directions is called **space quantization**. Like many aspects of quantum mechanics, this quantization of direction is totally unexpected. On the macroscopic scale, orbital angular momentum, such as that of the moon around the earth, can have any magnitude and be in any direction.

Detailed treatment of space quantization began to explain some complexities of atomic spectra, but certain patterns seemed to be caused by something else. As mentioned, spectral lines are actually closely spaced doublets, a characteristic called **fine structure**, as shown in [\[link\]](#). The doublet changes when a magnetic field is applied, implying that whatever causes the doublet interacts with a magnetic field. In 1925, Sem Goudsmit and George Uhlenbeck, two Dutch physicists, successfully argued that electrons have properties analogous to a macroscopic charge spinning on its axis. Electrons, in fact, have an internal or intrinsic angular momentum called **intrinsic spin \mathbf{S}** . Since electrons are charged, their intrinsic spin creates an **intrinsic magnetic field \mathbf{B}_{int}** , which interacts with their orbital magnetic field \mathbf{B}_{orb} . Furthermore, *electron intrinsic spin is quantized in magnitude and direction*, analogous to the situation for orbital angular momentum. The spin of the electron can have only one magnitude, and its direction can be at only one of two angles relative to a magnetic field, as seen in [\[link\]](#). We refer to this as spin up or spin down for the electron. Each spin direction has a different energy; hence, spectroscopic lines are split into two. Spectral doublets are now understood as being due to electron spin.



Fine structure. Upon close examination, spectral lines are doublets, even in the absence of an external magnetic field. The electron has an intrinsic magnetic field that interacts with its orbital magnetic field.



The intrinsic magnetic field B_{int} of an electron is attributed to its spin, S , roughly pictured to be due to its charge spinning on its axis. This is only a crude model, since electrons seem to have no size. The spin and intrinsic magnetic field of the electron can make only one of two angles with another magnetic field, such as that created by the electron's orbital

motion. Space is quantized for spin as well as for orbital angular momentum.

These two new insights—that the direction of angular momentum, whether orbital or spin, is quantized, and that electrons have intrinsic spin—help to explain many of the complexities of atomic and molecular spectra. In magnetic resonance imaging, it is the way that the intrinsic magnetic field of hydrogen and biological atoms interact with an external field that underlies the diagnostic fundamentals.

Section Summary

- The Zeeman effect—the splitting of lines when a magnetic field is applied—is caused by other quantized entities in atoms.
- Both the magnitude and direction of orbital angular momentum are quantized.
- The same is true for the magnitude and direction of the intrinsic spin of electrons.

Conceptual Questions

Exercise:

Problem:

What is the Zeeman effect, and what type of quantization was discovered because of this effect?

Glossary

Zeeman effect

the effect of external magnetic fields on spectral lines

intrinsic spin

the internal or intrinsic angular momentum of electrons

orbital angular momentum

an angular momentum that corresponds to the quantum analog of classical angular momentum

fine structure

the splitting of spectral lines of the hydrogen spectrum when the spectral lines are examined at very high resolution

space quantization

the fact that the orbital angular momentum can have only certain directions

intrinsic magnetic field

the magnetic field generated due to the intrinsic spin of electrons

orbital magnetic field

the magnetic field generated due to the orbital motion of electrons

Quantum Numbers and Rules

- Define quantum number.
- Calculate angle of angular momentum vector with an axis.
- Define spin quantum number.

Physical characteristics that are quantized—such as energy, charge, and angular momentum—are of such importance that names and symbols are given to them. The values of quantized entities are expressed in terms of **quantum numbers**, and the rules governing them are of the utmost importance in determining what nature is and does. This section covers some of the more important quantum numbers and rules—all of which apply in chemistry, material science, and far beyond the realm of atomic physics, where they were first discovered. Once again, we see how physics makes discoveries which enable other fields to grow.

The *energy states of bound systems are quantized*, because the particle wavelength can fit into the bounds of the system in only certain ways. This was elaborated for the hydrogen atom, for which the allowed energies are expressed as $E_n \propto 1/n^2$, where $n = 1, 2, 3, \dots$. We define n to be the principal quantum number that labels the basic states of a system. The lowest-energy state has $n = 1$, the first excited state has $n = 2$, and so on. Thus the allowed values for the principal quantum number are

Equation:

$$n = 1, 2, 3, \dots$$

This is more than just a numbering scheme, since the energy of the system, such as the hydrogen atom, can be expressed as some function of n , as can other characteristics (such as the orbital radii of the hydrogen atom).

The fact that the *magnitude of angular momentum is quantized* was first recognized by Bohr in relation to the hydrogen atom; it is now known to be true in general. With the development of quantum mechanics, it was found that the magnitude of angular momentum L can have only the values

Equation:

$$L = \sqrt{l(l+1)} \frac{h}{2\pi} \quad (l = 0, 1, 2, \dots, n-1),$$

where l is defined to be the **angular momentum quantum number**. The rule for l in atoms is given in the parentheses. Given n , the value of l can be any integer from zero up to $n - 1$. For example, if $n = 4$, then l can be 0, 1, 2, or 3.

Note that for $n = 1$, l can only be zero. This means that the ground-state angular momentum for hydrogen is actually zero, not $h/2\pi$ as Bohr proposed. The picture of circular orbits is not valid, because there would be angular momentum for any circular orbit. A more valid picture is the cloud of probability shown for the ground state of hydrogen in [\[link\]](#). The electron actually spends time in and near the nucleus. The reason the electron does not remain in the nucleus is related to Heisenberg's uncertainty principle—the electron's energy would have to be much too large to be confined to the small space of the nucleus. Now the first excited state of hydrogen has $n = 2$, so that l can be either 0 or 1, according to the rule in $L = \sqrt{l(l+1)} \frac{h}{2\pi}$. Similarly, for $n = 3$, l can be 0, 1, or 2. It is often most convenient to state the value of l , a simple integer, rather than calculating the value of L from $L = \sqrt{l(l+1)} \frac{h}{2\pi}$. For example, for $l = 2$, we see that

Equation:

$$L = \sqrt{2(2+1)} \frac{h}{2\pi} = \sqrt{6} \frac{h}{2\pi} = 0.390h = 2.58 \times 10^{-34} \text{ J} \cdot \text{s}.$$

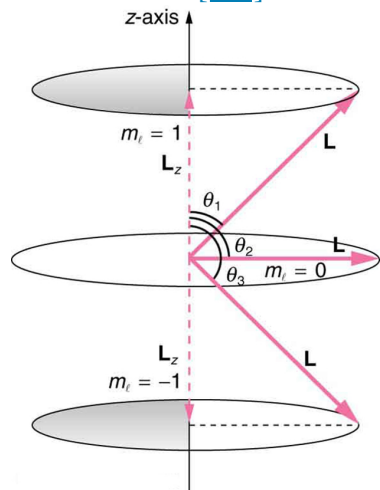
It is much simpler to state $l = 2$.

As recognized in the Zeeman effect, the *direction of angular momentum is quantized*. We now know this is true in all circumstances. It is found that the component of angular momentum along one direction in space, usually called the z -axis, can have only certain values of L_z . The direction in space must be related to something physical, such as the direction of the magnetic field at that location. This is an aspect of relativity. Direction has no meaning if there is nothing that varies with direction, as does magnetic force. The allowed values of L_z are

Equation:

$$L_z = m_l \frac{h}{2\pi} \quad (m_l = -l, -l+1, \dots, -1, 0, 1, \dots, l-1, l),$$

where L_z is the z -**component of the angular momentum** and m_l is the angular momentum projection quantum number. The rule in parentheses for the values of m_l is that it can range from $-l$ to l in steps of one. For example, if $l = 2$, then m_l can have the five values $-2, -1, 0, 1$, and 2 . Each m_l corresponds to a different energy in the presence of a magnetic field, so that they are related to the splitting of spectral lines into discrete parts, as discussed in the preceding section. If the z -component of angular momentum can have only certain values, then the angular momentum can have only certain directions, as illustrated in [\[link\]](#).



The component of a given angular momentum along the z -axis (defined by the direction of a magnetic field) can have only certain values; these are shown here for $l = 1$, for which $m_l = -1, 0$, and $+1$.

The direction of L is quantized in the sense that it can have only certain angles relative to the z -axis.

Example:

What Are the Allowed Directions?

Calculate the angles that the angular momentum vector \mathbf{L} can make with the z -axis for $l = 1$, as illustrated in [\[link\]](#).

Strategy

[\[link\]](#) represents the vectors \mathbf{L} and \mathbf{L}_z as usual, with arrows proportional to their magnitudes and pointing in the correct directions. \mathbf{L} and \mathbf{L}_z form a right triangle, with \mathbf{L} being the hypotenuse and \mathbf{L}_z the adjacent side. This means that the ratio of \mathbf{L}_z to \mathbf{L} is the cosine of the angle of interest. We can find \mathbf{L} and \mathbf{L}_z using $L = \sqrt{l(l+1)}\frac{h}{2\pi}$ and $L_z = m_l\frac{h}{2\pi}$.

Solution

We are given $l = 1$, so that m_l can be $+1$, 0 , or -1 . Thus L has the value given by $L = \sqrt{l(l+1)}\frac{h}{2\pi}$.

Equation:

$$L = \frac{\sqrt{l(l+1)}h}{2\pi} = \frac{\sqrt{2}h}{2\pi}$$

L_z can have three values, given by $L_z = m_l\frac{h}{2\pi}$.

Equation:

$$L_z = m_l\frac{h}{2\pi} = \begin{matrix} \frac{h}{2\pi}, & m_l = +1 \\ 0, & m_l = 0 \\ -\frac{h}{2\pi}, & m_l = -1 \end{matrix}$$

As can be seen in [\[link\]](#), $\cos \theta = L_z/L$, and so for $m_l = +1$, we have

Equation:

$$\cos \theta_1 = \frac{L_z}{L} = \frac{\frac{h}{2\pi}}{\frac{\sqrt{2}h}{2\pi}} = \frac{1}{\sqrt{2}} = 0.707.$$

Thus,

Equation:

$$\theta_1 = \cos^{-1}0.707 = 45.0^\circ.$$

Similarly, for $m_l = 0$, we find $\cos \theta_2 = 0$; thus,

Equation:

$$\theta_2 = \cos^{-1}0 = 90.0^\circ.$$

And for $m_l = -1$,

Equation:

$$\cos \theta_3 = \frac{L_z}{L} = \frac{-\frac{h}{2\pi}}{\frac{\sqrt{2}h}{2\pi}} = -\frac{1}{\sqrt{2}} = -0.707,$$

so that

Equation:

$$\theta_3 = \cos^{-1}(-0.707) = 135.0^\circ.$$

Discussion

The angles are consistent with the figure. Only the angle relative to the z -axis is quantized. L can point in any direction as long as it makes the proper angle with the z -axis. Thus the angular momentum vectors lie on cones as illustrated. This behavior is not observed on the large scale. To see how the correspondence principle holds here, consider that the smallest angle (θ_1 in the example) is for the maximum value of $m_l = 0$, namely $m_l = l$. For that smallest angle,

Equation:

$$\cos \theta = \frac{L_z}{L} = \frac{l}{\sqrt{l(l+1)}},$$

which approaches 1 as l becomes very large. If $\cos \theta = 1$, then $\theta = 0^\circ$. Furthermore, for large l , there are many values of m_l , so that all angles become possible as l gets very large.

Intrinsic Spin Angular Momentum Is Quantized in Magnitude and Direction

There are two more quantum numbers of immediate concern. Both were first discovered for electrons in conjunction with fine structure in atomic spectra. It is now well established that electrons and other fundamental particles have *intrinsic spin*, roughly analogous to a planet spinning on its axis. This spin is a fundamental characteristic of particles, and only one magnitude of intrinsic spin is allowed for a given type of particle. Intrinsic angular momentum is quantized independently of orbital angular momentum. Additionally, the direction of the spin is also quantized. It has been found that the **magnitude of the intrinsic (internal) spin angular momentum**, S , of an electron is given by

Equation:

$$S = \sqrt{s(s+1)} \frac{h}{2\pi} \quad (s = 1/2 \text{ for electrons}),$$

where s is defined to be the **spin quantum number**. This is very similar to the quantization of L given in $L = \sqrt{l(l+1)} \frac{h}{2\pi}$, except that the only value allowed for s for electrons is $1/2$.

The *direction of intrinsic spin is quantized*, just as is the direction of orbital angular momentum. The direction of spin angular momentum along one direction in space, again called the z -axis, can have only the values

Equation:

$$S_z = m_s \frac{h}{2\pi} \quad \left(m_s = -\frac{1}{2}, +\frac{1}{2} \right)$$

for electrons. S_z is the **z -component of spin angular momentum** and m_s is the **spin projection quantum number**. For electrons, s can only be $1/2$, and m_s can be either $+1/2$ or $-1/2$. Spin projection $m_s = +1/2$ is referred to as *spin up*, whereas $m_s = -1/2$ is called *spin down*. These are illustrated in [\[link\]](#).

Note:

Intrinsic Spin

In later chapters, we will see that intrinsic spin is a characteristic of all subatomic particles. For some particles s is half-integral, whereas for others s is integral—there are crucial differences between half-integral spin particles and integral spin particles. Protons and neutrons, like electrons, have $s = 1/2$, whereas photons have $s = 1$, and other particles called pions have $s = 0$, and so on.

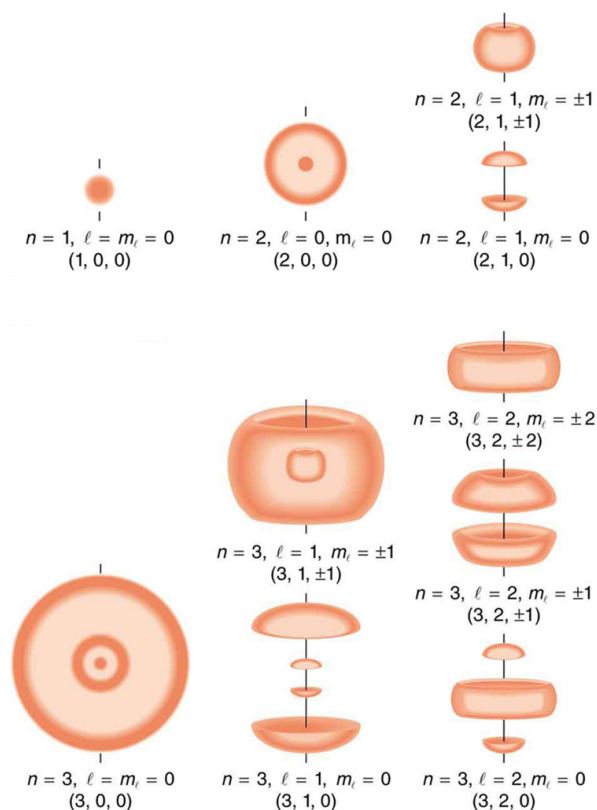
To summarize, the state of a system, such as the precise nature of an electron in an atom, is determined by its particular quantum numbers. These are expressed in the form (n, l, m_l, m_s) —see [\[link\]](#) For *electrons in atoms*, the principal quantum number can have the values $n = 1, 2, 3, \dots$. Once n is known, the values of the angular momentum quantum number are limited to $l = 1, 2, 3, \dots, n - 1$. For a given value of l , the angular momentum projection quantum number can have only the values $m_l = -l, -l + 1, \dots, -1, 0, 1, \dots, l - 1, l$. Electron spin is independent of n, l , and m_l , always having $s = 1/2$. The spin projection quantum number can have two values, $m_s = 1/2$ or $-1/2$.

Name	Symbol	Allowed values
Principal quantum number	n	1, 2, 3, ...
Angular momentum	l	0, 1, 2, ... $n - 1$
Angular momentum projection	m_l	$-l, -l + 1, \dots, -1, 0, 1, \dots, l - 1, l$ (or 0, $\pm 1, \pm 2, \dots, \pm l$)

Name	Symbol	Allowed values
Spin ^{footnote} The spin quantum number s is usually not stated, since it is always $1/2$ for electrons	s	$1/2$ (electrons)
Spin projection	m_s	$-1/2, +1/2$

Atomic Quantum Numbers

[\[link\]](#) shows several hydrogen states corresponding to different sets of quantum numbers. Note that these clouds of probability are the locations of electrons as determined by making repeated measurements—each measurement finds the electron in a definite location, with a greater chance of finding the electron in some places rather than others. With repeated measurements, the pattern of probability shown in the figure emerges. The clouds of probability do not look like nor do they correspond to classical orbits. The uncertainty principle actually prevents us and nature from knowing how the electron gets from one place to another, and so an orbit really does not exist as such. Nature on a small scale is again much different from that on the large scale.



Probability clouds for the electron in the ground state and several excited states of hydrogen. The nature of these states is determined by their sets of quantum numbers, here given as (n, l, m_l) . The ground state is (0, 0, 0); one of the possibilities for the second excited state is (3, 2, 1). The probability of finding the electron is indicated by the shade of color; the darker the coloring the greater the chance of finding the electron.

We will see that the quantum numbers discussed in this section are valid for a broad range of particles and other systems, such as nuclei. Some quantum numbers, such as intrinsic spin, are related to fundamental classifications of subatomic particles, and they obey laws that will give us further insight into the substructure of matter and its interactions.

Note:

PhET Explorations: Stern-Gerlach Experiment

The classic Stern-Gerlach Experiment shows that atoms have a property called spin. Spin is a kind of intrinsic angular momentum, which has no classical counterpart. When the z-component of the spin is

measured, one always gets one of two values: spin up or spin down.

https://phet.colorado.edu/sims/stern-gerlach/stern-gerlach_en.html

Section Summary

- Quantum numbers are used to express the allowed values of quantized entities. The principal quantum number n labels the basic states of a system and is given by

Equation:

$$n = 1, 2, 3, \dots$$

- The magnitude of angular momentum is given by

Equation:

$$L = \sqrt{l(l+1)} \frac{h}{2\pi} \quad (l = 0, 1, 2, \dots, n-1),$$

where l is the angular momentum quantum number. The direction of angular momentum is quantized, in that its component along an axis defined by a magnetic field, called the z -axis is given by

Equation:

$$L_z = m_l \frac{h}{2\pi} \quad (m_l = -l, -l+1, \dots, -1, 0, 1, \dots, l-1, l),$$

where L_z is the z -component of the angular momentum and m_l is the angular momentum projection quantum number. Similarly, the electron's intrinsic spin angular momentum S is given by

Equation:

$$S = \sqrt{s(s+1)} \frac{h}{2\pi} \quad (s = 1/2 \text{ for electrons}),$$

s is defined to be the spin quantum number. Finally, the direction of the electron's spin along the z -axis is given by

Equation:

$$S_z = m_s \frac{h}{2\pi} \quad \left(m_s = -\frac{1}{2}, +\frac{1}{2} \right),$$

where S_z is the z -component of spin angular momentum and m_s is the spin projection quantum number. Spin projection $m_s = +1/2$ is referred to as spin up, whereas $m_s = -1/2$ is called spin down. [\[link\]](#) summarizes the atomic quantum numbers and their allowed values.

Conceptual Questions

Exercise:

Problem: Define the quantum numbers n , l , m_l , s , and m_s .

Exercise:

Problem: For a given value of n , what are the allowed values of l ?

Exercise:

Problem:

For a given value of l , what are the allowed values of m_l ? What are the allowed values of m_l for a given value of n ? Give an example in each case.

Exercise:

Problem:

List all the possible values of s and m_s for an electron. Are there particles for which these values are different? The same?

Problem Exercises

Exercise:

Problem:

If an atom has an electron in the $n = 5$ state with $m_l = 3$, what are the possible values of l ?

Solution:

$l = 4, 3$ are possible since $l < n$ and $|m_l| \leq l$.

Exercise:

Problem: An atom has an electron with $m_l = 2$. What is the smallest value of n for this electron?

Exercise:

Problem: What are the possible values of m_l for an electron in the $n = 4$ state?

Solution:

$n = 4 \Rightarrow l = 3, 2, 1, 0 \Rightarrow m_l = \pm 3, \pm 2, \pm 1, 0$ are possible.

Exercise:

Problem:

What, if any, constraints does a value of $m_l = 1$ place on the other quantum numbers for an electron in an atom?

Exercise:

Problem:

(a) Calculate the magnitude of the angular momentum for an $l = 1$ electron. (b) Compare your answer to the value Bohr proposed for the $n = 1$ state.

Solution:

(a) $1.49 \times 10^{-34} \text{ J} \cdot \text{s}$

(b) $1.06 \times 10^{-34} \text{ J} \cdot \text{s}$

Exercise:**Problem:**

(a) What is the magnitude of the angular momentum for an $l = 1$ electron? (b) Calculate the magnitude of the electron's spin angular momentum. (c) What is the ratio of these angular momenta?

Exercise:

Problem: Repeat [\[link\]](#) for $l = 3$.

Solution:

(a) $3.66 \times 10^{-34} \text{ J} \cdot \text{s}$

(b) $s = 9.13 \times 10^{-35} \text{ J} \cdot \text{s}$

(c) $\frac{L}{s} = \frac{\sqrt{12}}{\sqrt{3/4}} = 4$

Exercise:**Problem:**

(a) How many angles can L make with the z -axis for an $l = 2$ electron? (b) Calculate the value of the smallest angle.

Exercise:

Problem: What angles can the spin S of an electron make with the z -axis?

Solution:

$$\theta = 54.7^\circ, 125.3^\circ$$

Glossary

quantum numbers

the values of quantized entities, such as energy and angular momentum

angular momentum quantum number

a quantum number associated with the angular momentum of electrons

spin quantum number

the quantum number that parameterizes the intrinsic angular momentum (or spin angular momentum, or simply spin) of a given particle

spin projection quantum number

quantum number that can be used to calculate the intrinsic electron angular momentum along the z -axis

z -component of spin angular momentum

component of intrinsic electron spin along the z -axis

magnitude of the intrinsic (internal) spin angular momentum

given by $S = \sqrt{s(s+1)} \frac{h}{2\pi}$

z -component of the angular momentum

component of orbital angular momentum of electron along the z -axis

The Pauli Exclusion Principle

- Define the composition of an atom along with its electrons, neutrons, and protons.
- Explain the Pauli exclusion principle and its application to the atom.
- Specify the shell and subshell symbols and their positions.
- Define the position of electrons in different shells of an atom.
- State the position of each element in the periodic table according to shell filling.

Multiple-Electron Atoms

All atoms except hydrogen are multiple-electron atoms. The physical and chemical properties of elements are directly related to the number of electrons a neutral atom has. The periodic table of the elements groups elements with similar properties into columns. This systematic organization is related to the number of electrons in a neutral atom, called the **atomic number**, Z . We shall see in this section that the exclusion principle is key to the underlying explanations, and that it applies far beyond the realm of atomic physics.

In 1925, the Austrian physicist Wolfgang Pauli (see [link](#)) proposed the following rule: No two electrons can have the same set of quantum numbers. That is, no two electrons can be in the same state. This statement is known as the **Pauli exclusion principle**, because it excludes electrons from being in the same state. The Pauli exclusion principle is extremely powerful and very broadly applicable. It applies to any identical particles with half-integral intrinsic spin—that is, having $s = 1/2, 3/2, \dots$. Thus no two electrons can have the same set of quantum numbers.

Note:

Pauli Exclusion Principle

No two electrons can have the same set of quantum numbers. That is, no two electrons can be in the same state.

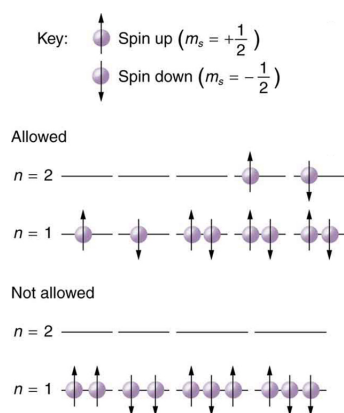


The Austrian physicist Wolfgang Pauli (1900–1958) played a major role in the development of quantum mechanics. He proposed the exclusion principle; hypothesized the existence of an important particle,

called the neutrino,
before it was directly
observed; made
fundamental
contributions to
several areas of
theoretical physics;
and influenced many
students who went
on to do important
work of their own.
(credit: Nobel
Foundation, via
Wikimedia
Commons)

Let us examine how the exclusion principle applies to electrons in atoms. The quantum numbers involved were defined in [Quantum Numbers and Rules](#) as n , l , m_l , s , and m_s . Since s is always $1/2$ for electrons, it is redundant to list s , and so we omit it and specify the state of an electron by a set of four numbers (n, l, m_l, m_s) . For example, the quantum numbers $(2, 1, 0, -1/2)$ completely specify the state of an electron in an atom.

Since no two electrons can have the same set of quantum numbers, there are limits to how many of them can be in the same energy state. Note that n determines the energy state in the absence of a magnetic field. So we first choose n , and then we see how many electrons can be in this energy state or energy level. Consider the $n = 1$ level, for example. The only value l can have is 0 (see [link](#) for a list of possible values once n is known), and thus m_l can only be 0. The spin projection m_s can be either $+1/2$ or $-1/2$, and so there can be two electrons in the $n = 1$ state. One has quantum numbers $(1, 0, 0, +1/2)$, and the other has $(1, 0, 0, -1/2)$. [link](#) illustrates that there can be one or two electrons having $n = 1$, but not three.



The Pauli exclusion
principle explains why
some configurations of
electrons are allowed
while others are not.
Since electrons cannot
have the same set of
quantum numbers, a
maximum of two can be

in the $n = 1$ level, and a third electron must reside in the higher-energy $n = 2$ level. If there are two electrons in the $n = 1$ level, their spins must be in opposite directions. (More precisely, their spin projections must differ.)

Shells and Subshells

Because of the Pauli exclusion principle, only hydrogen and helium can have all of their electrons in the $n = 1$ state. Lithium (see the periodic table) has three electrons, and so one must be in the $n = 2$ level. This leads to the concept of shells and shell filling. As we progress up in the number of electrons, we go from hydrogen to helium, lithium, beryllium, boron, and so on, and we see that there are limits to the number of electrons for each value of n . Higher values of the shell n correspond to higher energies, and they can allow more electrons because of the various combinations of l , m_l , and m_s that are possible. Each value of the principal quantum number n thus corresponds to an atomic **shell** into which a limited number of electrons can go. Shells and the number of electrons in them determine the physical and chemical properties of atoms, since it is the outermost electrons that interact most with anything outside the atom.

The probability clouds of electrons with the lowest value of l are closest to the nucleus and, thus, more tightly bound. Thus when shells fill, they start with $l = 0$, progress to $l = 1$, and so on. Each value of l thus corresponds to a **subshell**.

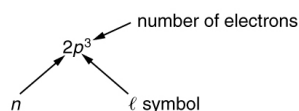
The table given below lists symbols traditionally used to denote shells and subshells.

Shell	Subshell	
n	l	Symbol
1	0	s
2	1	p
3	2	d
4	3	f

Shell	Subshell	
5	4	<i>g</i>
	5	<i>h</i>
	6 ^[footnote] It is unusual to deal with subshells having <i>l</i> greater than 6, but when encountered, they continue to be labeled in alphabetical order.	<i>i</i>

Shell and Subshell Symbols

To denote shells and subshells, we write nl with a number for n and a letter for l . For example, an electron in the $n = 1$ state must have $l = 0$, and it is denoted as a $1s$ electron. Two electrons in the $n = 1$ state is denoted as $1s^2$. Another example is an electron in the $n = 2$ state with $l = 1$, written as $2p$. The case of three electrons with these quantum numbers is written $2p^3$. This notation, called spectroscopic notation, is generalized as shown in [\[link\]](#).



Counting the number of possible combinations of quantum numbers allowed by the exclusion principle, we can determine how many electrons it takes to fill each subshell and shell.

Example:

How Many Electrons Can Be in This Shell?

List all the possible sets of quantum numbers for the $n = 2$ shell, and determine the number of electrons that can be in the shell and each of its subshells.

Strategy

Given $n = 2$ for the shell, the rules for quantum numbers limit l to be 0 or 1. The shell therefore has two subshells, labeled $2s$ and $2p$. Since the lowest l subshell fills first, we start with the $2s$ subshell possibilities and then proceed with the $2p$ subshell.

Solution

It is convenient to list the possible quantum numbers in a table, as shown below.

n	ℓ	m_ℓ	m_s	Subshell	Total in subshell	Total in shell
2	0	0	+1/2	2s	2	8
2	0	0	-1/2			
2	1	1	+1/2	2p	6	
2	1	1	-1/2			
2	1	0	+1/2			
2	1	0	-1/2			
2	1	-1	+1/2			
2	1	-1	-1/2			

Discussion

It is laborious to make a table like this every time we want to know how many electrons can be in a shell or subshell. There exist general rules that are easy to apply, as we shall now see.

The number of electrons that can be in a subshell depends entirely on the value of l . Once l is known, there are a fixed number of values of m_l , each of which can have two values for m_s . First, since m_l goes from $-l$ to l in steps of 1, there are $2l + 1$ possibilities. This number is multiplied by 2, since each electron can be spin up or spin down. Thus the *maximum number of electrons that can be in a subshell* is $2(2l + 1)$.

For example, the $2s$ subshell in [\[link\]](#) has a maximum of 2 electrons in it, since $2(2l + 1) = 2(0 + 1) = 2$ for this subshell. Similarly, the $2p$ subshell has a maximum of 6 electrons, since $2(2l + 1) = 2(2 + 1) = 6$. For a shell, the maximum number is the sum of what can fit in the subshells. Some algebra shows that the *maximum number of electrons that can be in a shell* is $2n^2$.

For example, for the first shell $n = 1$, and so $2n^2 = 2$. We have already seen that only two electrons can be in the $n = 1$ shell. Similarly, for the second shell, $n = 2$, and so $2n^2 = 8$. As found in [\[link\]](#), the total number of electrons in the $n = 2$ shell is 8.

Example:

Subshells and Totals for $n = 3$

How many subshells are in the $n = 3$ shell? Identify each subshell, calculate the maximum number of electrons that will fit into each, and verify that the total is $2n^2$.

Strategy

Subshells are determined by the value of l ; thus, we first determine which values of l are allowed, and then we apply the equation “maximum number of electrons that can be in a subshell = $2(2l + 1)$ ” to find the number of electrons in each subshell.

Solution

Since $n = 3$, we know that l can be 0, 1, or 2; thus, there are three possible subshells. In standard notation, they are labeled the $3s$, $3p$, and $3d$ subshells. We have already seen that 2 electrons can be in an s state, and 6 in a p state, but let us use the equation “maximum number of electrons that can be in a subshell = $2(2l + 1)$ ” to calculate the maximum number in each:

Equation:

$$\begin{aligned} 3s \text{ has } l = 0; \text{ thus, } 2(2l + 1) &= 2(0 + 1) = 2 \\ 3p \text{ has } l = 1; \text{ thus, } 2(2l + 1) &= 2(2 + 1) = 6 \\ 3d \text{ has } l = 2; \text{ thus, } 2(2l + 1) &= 2(4 + 1) = 10 \\ \text{Total} &= 18 \\ &(\text{in the } n = 3 \text{ shell}) \end{aligned}$$

The equation “maximum number of electrons that can be in a shell = $2n^2$ ” gives the maximum number in the $n = 3$ shell to be

Equation:

$$\text{Maximum number of electrons} = 2n^2 = 2(3)^2 = 2(9) = 18.$$

Discussion

The total number of electrons in the three possible subshells is thus the same as the formula $2n^2$. In standard (spectroscopic) notation, a filled $n = 3$ shell is denoted as $3s^23p^63d^{10}$. Shells do not fill in a simple manner. Before the $n = 3$ shell is completely filled, for example, we begin to find electrons in the $n = 4$ shell.

Shell Filling and the Periodic Table

[\[link\]](#) shows electron configurations for the first 20 elements in the periodic table, starting with hydrogen and its single electron and ending with calcium. The Pauli exclusion principle determines the maximum number of electrons allowed in each shell and subshell. But the order in which the shells and subshells are filled is complicated because of the large numbers of interactions between electrons.

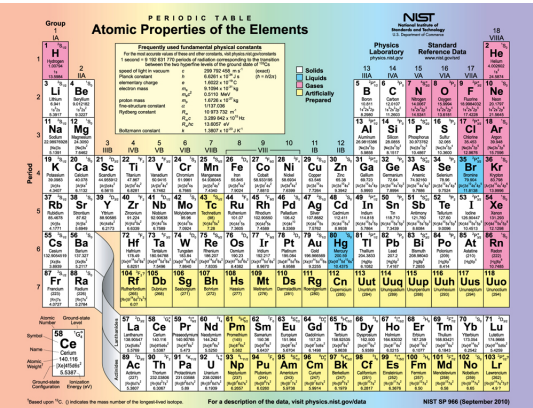
Element	Number of electrons (Z)	Ground state configuration					
H	1	$1s^1$					
He	2	$1s^2$					
Li	3	$1s^2$	$2s^1$				
Be	4	"	$2s^2$				
B	5	"	$2s^2$	$2p^1$			
C	6	"	$2s^2$	$2p^2$			
N	7	"	$2s^2$	$2p^3$			
O	8	"	$2s^2$	$2p^4$			
F	9	"	$2s^2$	$2p^5$			
Ne	10	"	$2s^2$	$2p^6$			
Na	11	"	$2s^2$	$2p^6$	$3s^1$		
Mg	12	"	"	"	$3s^2$		
Al	13	"	"	"	$3s^2$	$3p^1$	
Si	14	"	"	"	$3s^2$	$3p^2$	

Element	Number of electrons (Z)	Ground state configuration					
P	15	"	"	"	$3s^2$	$3p^3$	
S	16	"	"	"	$3s^2$	$3p^4$	
Cl	17	"	"	"	$3s^2$	$3p^5$	
Ar	18	"	"	"	$3s^2$	$3p^6$	
K	19	"	"	"	$3s^2$	$3p^6$	$4s^1$
Ca	20	"	"	"	"	"	$4s^2$

Electron Configurations of Elements Hydrogen Through Calcium

Examining the above table, you can see that as the number of electrons in an atom increases from 1 in hydrogen to 2 in helium and so on, the lowest-energy shell gets filled first—that is, the $n = 1$ shell fills first, and then the $n = 2$ shell begins to fill. Within a shell, the subshells fill starting with the lowest l , or with the s subshell, then the p , and so on, usually until all subshells are filled. The first exception to this occurs for potassium, where the $4s$ subshell begins to fill before any electrons go into the $3d$ subshell. The next exception is not shown in [\[link\]](#); it occurs for rubidium, where the $5s$ subshell starts to fill before the $4d$ subshell. The reason for these exceptions is that $l = 0$ electrons have probability clouds that penetrate closer to the nucleus and, thus, are more tightly bound (lower in energy).

[\[link\]](#) shows the periodic table of the elements, through element 118. Of special interest are elements in the main groups, namely, those in the columns numbered 1, 2, 13, 14, 15, 16, 17, and 18.



Periodic table of the elements (credit:

The number of electrons in the outermost subshell determines the atom's chemical properties, since it is these electrons that are farthest from the nucleus and thus interact most with other atoms. If the outermost subshell can accept or give up an electron easily, then the atom will be highly reactive chemically. Each group in the periodic table is characterized by its outermost electron configuration. Perhaps the most familiar is Group 18 (Group VIII), the noble gases (helium, neon, argon, etc.). These gases are all characterized by a filled outer subshell that is particularly stable. This means that they have large ionization energies and do not readily give up an electron. Furthermore, if they were to accept an extra electron, it would be in a significantly higher level and thus loosely bound. Chemical reactions often involve sharing electrons. Noble gases can be forced into unstable chemical compounds only under high pressure and temperature.

Group 17 (Group VII) contains the halogens, such as fluorine, chlorine, iodine and bromine, each of which has one less electron than a neighboring noble gas. Each halogen has 5 p electrons (a p^5 configuration), while the p subshell can hold 6 electrons. This means the halogens have one vacancy in their outermost subshell. They thus readily accept an extra electron (it becomes tightly bound, closing the shell as in noble gases) and are highly reactive chemically. The halogens are also likely to form singly negative ions, such as Cl^- , fitting an extra electron into the vacancy in the outer subshell. In contrast, alkali metals, such as sodium and potassium, all have a single s electron in their outermost subshell (an s^1 configuration) and are members of Group 1 (Group I). These elements easily give up their extra electron and are thus highly reactive chemically. As you might expect, they also tend to form singly positive ions, such as Na^+ , by losing their loosely bound outermost electron. They are metals (conductors), because the loosely bound outer electron can move freely.

Of course, other groups are also of interest. Carbon, silicon, and germanium, for example, have similar chemistries and are in Group 4 (Group IV). Carbon, in particular, is extraordinary in its ability to form many types of bonds and to be part of long chains, such as inorganic molecules. The large group of what are called transitional elements is characterized by the filling of the d subshells and crossing of energy levels. Heavier groups, such as the lanthanide series, are more complex—their shells do not fill in simple order. But the groups recognized by chemists such as Mendeleev have an explanation in the substructure of atoms.

Note:

PhET Explorations: Stern-Gerlach Experiment

Build an atom out of protons, neutrons, and electrons, and see how the element, charge, and mass change. Then play a game to test your ideas!

https://phet.colorado.edu/sims/html/build-an-atom/latest/build-an-atom_en.html

Section Summary

- The state of a system is completely described by a complete set of quantum numbers. This set is written as (n, l, m_l, m_s) .
- The Pauli exclusion principle says that no two electrons can have the same set of quantum numbers; that is, no two electrons can be in the same state.
- This exclusion limits the number of electrons in atomic shells and subshells. Each value of n corresponds to a shell, and each value of l corresponds to a subshell.
- The maximum number of electrons that can be in a subshell is $2(2l + 1)$.
- The maximum number of electrons that can be in a shell is $2n^2$.

Conceptual Questions

Exercise:**Problem:**

Identify the shell, subshell, and number of electrons for the following: (a) $2p^3$. (b) $4d^9$. (c) $3s^1$. (d) $5g^{16}$.

Exercise:**Problem:**

Which of the following are not allowed? State which rule is violated for any that are not allowed. (a) $1p^3$ (b) $2p^8$ (c) $3g^{11}$ (d) $4f^2$

Problem Exercises**Exercise:**

Problem: (a) How many electrons can be in the $n = 4$ shell?

(b) What are its subshells, and how many electrons can be in each?

Solution:

(a) 32. (b) 2 in s , 6 in p , 10 in d , and 14 in f , for a total of 32.

Exercise:

Problem: (a) What is the minimum value of l for a subshell that has 11 electrons in it?

(b) If this subshell is in the $n = 5$ shell, what is the spectroscopic notation for this atom?

Exercise:**Problem:**

(a) If one subshell of an atom has 9 electrons in it, what is the minimum value of l ? (b) What is the spectroscopic notation for this atom, if this subshell is part of the $n = 3$ shell?

Solution:

(a) 2

(b) $3d^9$

Exercise:**Problem:**

(a) List all possible sets of quantum numbers (n, l, m_l, m_s) for the $n = 3$ shell, and determine the number of electrons that can be in the shell and each of its subshells.

(b) Show that the number of electrons in the shell equals $2n^2$ and that the number in each subshell is $2(2l + 1)$.

Exercise:**Problem:**

Which of the following spectroscopic notations are not allowed? (a) $5s^1$ (b) $1d^1$ (c) $4s^3$ (d) $3p^7$ (e) $5g^{15}$. State which rule is violated for each that is not allowed.

Solution:

(b) $n \geq l$ is violated,

(c) cannot have 3 electrons in s subshell since $3 > (2l + 1) = 2$

(d) cannot have 7 electrons in p subshell since $7 > (2l + 1) = 2(2 + 1) = 6$

Exercise:**Problem:**

Which of the following spectroscopic notations are allowed (that is, which violate none of the rules regarding values of quantum numbers)? (a) $1s^1$ (b) $1d^3$ (c) $4s^2$ (d) $3p^7$ (e) $6h^{20}$

Exercise:**Problem:**

(a) Using the Pauli exclusion principle and the rules relating the allowed values of the quantum numbers (n, l, m_l, m_s), prove that the maximum number of electrons in a subshell is $2n^2$.

(b) In a similar manner, prove that the maximum number of electrons in a shell is $2n^2$.

Solution:

(a) The number of different values of m_l is $\pm l, \pm (l - 1), \dots, 0$ for each $l > 0$ and one for $l = 0 \Rightarrow (2l + 1)$. Also an overall factor of 2 since each m_l can have m_s equal to either $+1/2$ or $-1/2 \Rightarrow 2(2l + 1)$.

(b) for each value of l , you get $2(2l + 1)$

$$= 0, 1, 2, \dots, (n-1) \Rightarrow 2\{[(2)(0) + 1] + [(2)(1) + 1] + \dots + [(2)(n-1) + 1]\} = 2[1 + 3 + \dots + (2n-3) +$$

n terms

to see that the expression in the box is $= n^2$, imagine taking $(n - 1)$ from the last term and adding it to first term $= 2[1 + (n-1) + 3 + \dots + (2n-3) + (2n-1) - (n-1)] = 2[n + 3 + \dots + (2n-3) + n]$. Now take $(n - 3)$ from penultimate term and add to the second term $2[n + n + \dots + n + n] = 2n^2$.

n terms

Exercise:**Problem: Integrated Concepts**

Estimate the density of a nucleus by calculating the density of a proton, taking it to be a sphere 1.2 fm in diameter. Compare your result with the value estimated in this chapter.

Exercise:**Problem: Integrated Concepts**

The electric and magnetic forces on an electron in the CRT in [\[link\]](#) are supposed to be in opposite directions. Verify this by determining the direction of each force for the situation shown. Explain how you obtain the directions (that is, identify the rules used).

Solution:

The electric force on the electron is up (toward the positively charged plate). The magnetic force is down (by the RHR).

Exercise:

Problem:

- (a) What is the distance between the slits of a diffraction grating that produces a first-order maximum for the first Balmer line at an angle of 20.0° ?
- (b) At what angle will the fourth line of the Balmer series appear in first order?
- (c) At what angle will the second-order maximum be for the first line?

Exercise:**Problem: Integrated Concepts**

A galaxy moving away from the earth has a speed of $0.0100c$. What wavelength do we observe for an $n_i = 7$ to $n_f = 2$ transition for hydrogen in that galaxy?

Solution:

401 nm

Exercise:**Problem: Integrated Concepts**

Calculate the velocity of a star moving relative to the earth if you observe a wavelength of 91.0 nm for ionized hydrogen capturing an electron directly into the lowest orbital (that is, a $n_i = \infty$ to $n_f = 1$, or a Lyman series transition).

Exercise:**Problem: Integrated Concepts**

In a Millikan oil-drop experiment using a setup like that in [\[link\]](#), a 500-V potential difference is applied to plates separated by 2.50 cm. (a) What is the mass of an oil drop having two extra electrons that is suspended motionless by the field between the plates? (b) What is the diameter of the drop, assuming it is a sphere with the density of olive oil?

Solution:

(a) 6.54×10^{-16} kg

(b) 5.54×10^{-7} m

Exercise:**Problem: Integrated Concepts**

What double-slit separation would produce a first-order maximum at 3.00° for 25.0-keV x rays? The small answer indicates that the wave character of x rays is best determined by having them interact with very small objects such as atoms and molecules.

Exercise:**Problem: Integrated Concepts**

In a laboratory experiment designed to duplicate Thomson's determination of q_e/m_e , a beam of electrons having a velocity of 6.00×10^7 m/s enters a 5.00×10^{-3} T magnetic field. The beam moves perpendicular

to the field in a path having a 6.80-cm radius of curvature. Determine q_e/m_e from these observations, and compare the result with the known value.

Solution:

$1.76 \times 10^{11} \text{ C/kg}$, which agrees with the known value of $1.759 \times 10^{11} \text{ C/kg}$ to within the precision of the measurement

Exercise:

Problem: Integrated Concepts

Find the value of l , the orbital angular momentum quantum number, for the moon around the earth. The extremely large value obtained implies that it is impossible to tell the difference between adjacent quantized orbits for macroscopic objects.

Exercise:

Problem: Integrated Concepts

Particles called muons exist in cosmic rays and can be created in particle accelerators. Muons are very similar to electrons, having the same charge and spin, but they have a mass 207 times greater. When muons are captured by an atom, they orbit just like an electron but with a smaller radius, since the mass in

$$a_B = \frac{h^2}{4\pi^2 m_e k q_e^2} = 0.529 \times 10^{-10} \text{ m is } 207 m_e.$$

(a) Calculate the radius of the $n = 1$ orbit for a muon in a uranium ion ($Z = 92$).

(b) Compare this with the 7.5-fm radius of a uranium nucleus. Note that since the muon orbits inside the electron, it falls into a hydrogen-like orbit. Since your answer is less than the radius of the nucleus, you can see that the photons emitted as the muon falls into its lowest orbit can give information about the nucleus.

Solution:

(a) 2.78 fm

(b) 0.37 of the nuclear radius.

Exercise:

Problem: Integrated Concepts

Calculate the minimum amount of energy in joules needed to create a population inversion in a helium-neon laser containing 1.00×10^{-4} moles of neon.

Exercise:

Problem: Integrated Concepts

A carbon dioxide laser used in surgery emits infrared radiation with a wavelength of 10.6 μm . In 1.00 ms, this laser raised the temperature of 1.00 cm^3 of flesh to 100°C and evaporated it.

(a) How many photons were required? You may assume flesh has the same heat of vaporization as water. (b) What was the minimum power output during the flash?

Solution:

(a) 1.34×10^{23}

(b) 2.52 MW

Exercise:

Problem: Integrated Concepts

Suppose an MRI scanner uses 100-MHz radio waves.

- (a) Calculate the photon energy.
- (b) How does this compare to typical molecular binding energies?

Exercise:

Problem: Integrated Concepts

- (a) An excimer laser used for vision correction emits 193-nm UV. Calculate the photon energy in eV.
- (b) These photons are used to evaporate corneal tissue, which is very similar to water in its properties. Calculate the amount of energy needed per molecule of water to make the phase change from liquid to gas. That is, divide the heat of vaporization in kJ/kg by the number of water molecules in a kilogram.
- (c) Convert this to eV and compare to the photon energy. Discuss the implications.

Solution:

- (a) 6.42 eV
- (b) 7.27×10^{-20} J/molecule
- (c) 0.454 eV, 14.1 times less than a single UV photon. Therefore, each photon will evaporate approximately 14 molecules of tissue. This gives the surgeon a rather precise method of removing corneal tissue from the surface of the eye.

Exercise:

Problem: Integrated Concepts

A neighboring galaxy rotates on its axis so that stars on one side move toward us as fast as 200 km/s, while those on the other side move away as fast as 200 km/s. This causes the EM radiation we receive to be Doppler shifted by velocities over the entire range of ± 200 km/s. What range of wavelengths will we observe for the 656.0-nm line in the Balmer series of hydrogen emitted by stars in this galaxy. (This is called line broadening.)

Exercise:

Problem: Integrated Concepts

A pulsar is a rapidly spinning remnant of a supernova. It rotates on its axis, sweeping hydrogen along with it so that hydrogen on one side moves toward us as fast as 50.0 km/s, while that on the other side moves away as fast as 50.0 km/s. This means that the EM radiation we receive will be Doppler shifted over a range of ± 50.0 km/s. What range of wavelengths will we observe for the 91.20-nm line in the Lyman series of hydrogen? (Such line broadening is observed and actually provides part of the evidence for rapid rotation.)

Solution:

91.18 nm to 91.22 nm

Exercise:

Problem: Integrated Concepts

Prove that the velocity of charged particles moving along a straight path through perpendicular electric and magnetic fields is $v = E/B$. Thus crossed electric and magnetic fields can be used as a velocity selector independent of the charge and mass of the particle involved.

Exercise:**Problem: Unreasonable Results**

(a) What voltage must be applied to an X-ray tube to obtain 0.0100-nm-wavelength X-rays for use in exploring the details of nuclei? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) 1.24×10^{11} V

(b) The voltage is extremely large compared with any practical value.

(c) The assumption of such a short wavelength by this method is unreasonable.

Exercise:**Problem: Unreasonable Results**

A student in a physics laboratory observes a hydrogen spectrum with a diffraction grating for the purpose of measuring the wavelengths of the emitted radiation. In the spectrum, she observes a yellow line and finds its wavelength to be 589 nm. (a) Assuming this is part of the Balmer series, determine n_i , the principal quantum number of the initial state. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:**Problem: Construct Your Own Problem**

The solar corona is so hot that most atoms in it are ionized. Consider a hydrogen-like atom in the corona that has only a single electron. Construct a problem in which you calculate selected spectral energies and wavelengths of the Lyman, Balmer, or other series of this atom that could be used to identify its presence in a very hot gas. You will need to choose the atomic number of the atom, identify the element, and choose which spectral lines to consider.

Exercise:**Problem: Construct Your Own Problem**

Consider the Doppler-shifted hydrogen spectrum received from a rapidly receding galaxy. Construct a problem in which you calculate the energies of selected spectral lines in the Balmer series and examine whether they can be described with a formula like that in the equation $\frac{1}{\lambda} = R \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right)$, but with a different constant R .

Glossary

atomic number

the number of protons in the nucleus of an atom

Pauli exclusion principle

a principle that states that no two electrons can have the same set of quantum numbers; that is, no two electrons can be in the same state

shell

a probability cloud for electrons that has a single principal quantum number

subshell

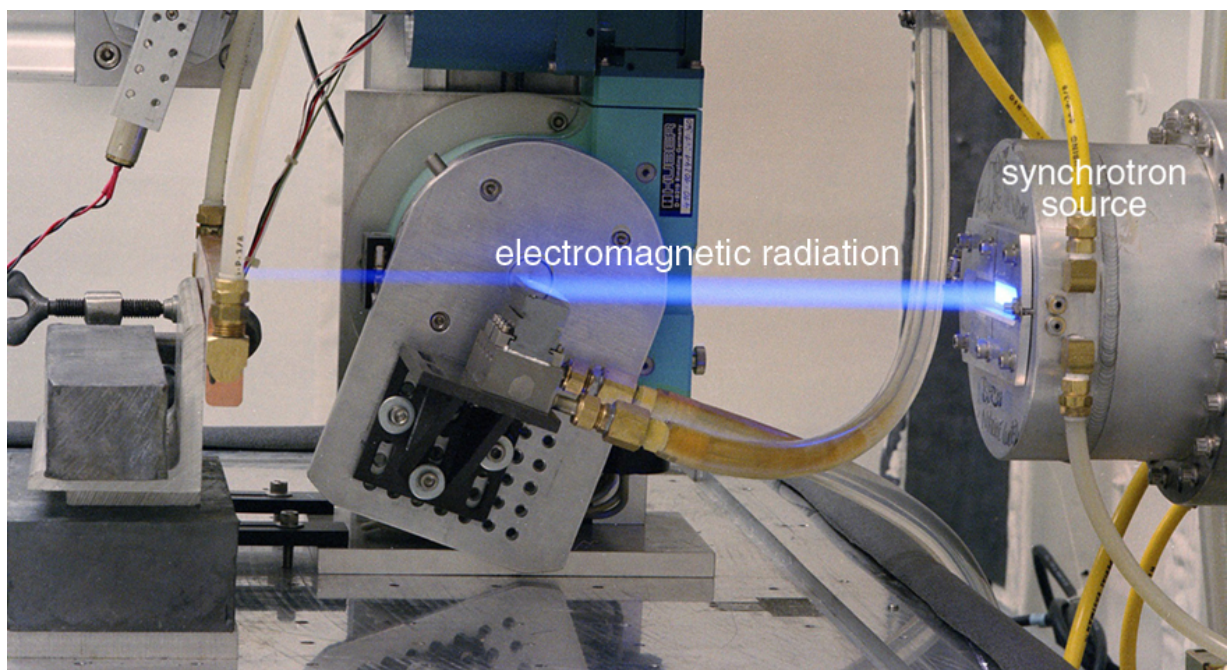
the probability cloud for electrons that has a single angular momentum quantum number l

Introduction to Radioactivity and Nuclear Physics

class="introduction"

- Define radioactivity.

The
synchrotron
source
produces
electromagnet
ic radiation, as
evident from
the visible
glow. (credit:
United States
Department of
Energy, via
Wikimedia
Commons)

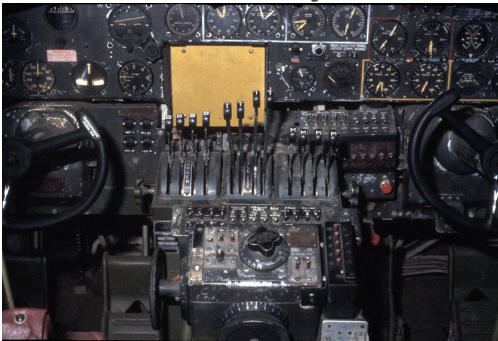


There is an ongoing quest to find substructures of matter. At one time, it was thought that atoms would be the ultimate substructure, but just when the first direct evidence of atoms was obtained, it became clear that they have a substructure and a tiny *nucleus*. The nucleus itself has spectacular characteristics. For example, certain nuclei are unstable, and their decay emits radiations with energies millions of times greater than atomic energies. Some of the mysteries of nature, such as why the core of the earth remains molten and how the sun produces its energy, are explained by nuclear phenomena. The exploration of *radioactivity* and the nucleus revealed fundamental and previously unknown particles, forces, and conservation laws. That exploration has evolved into a search for further underlying structures, such as quarks. In this chapter, the fundamentals of nuclear radioactivity and the nucleus are explored. The following two chapters explore the more important applications of nuclear physics in the field of medicine. We will also explore the basics of what we know about quarks and other substructures smaller than nuclei.

Nuclear Radioactivity

- Explain nuclear radiation.
- Explain the types of radiation—alpha emission, beta emission, and gamma emission.
- Explain the ionization of radiation in an atom.
- Define the range of radiation.

The discovery and study of nuclear radioactivity quickly revealed evidence of revolutionary new physics. In addition, uses for nuclear radiation also emerged quickly—for example, people such as Ernest Rutherford used it to determine the size of the nucleus and devices were painted with radon-doped paint to make them glow in the dark (see [\[link\]](#)). We therefore begin our study of nuclear physics with the discovery and basic features of nuclear radioactivity.



The dials of this World War II aircraft glow in the dark, because they are painted with radium-doped phosphorescent paint. It is a poignant reminder of the dual nature of radiation. Although radium paint dials are conveniently visible day and night, they emit radon, a radioactive gas that is hazardous and is not

directly sensed. (credit:
U.S. Air Force Photo)

Discovery of Nuclear Radioactivity

In 1896, the French physicist Antoine Henri Becquerel (1852–1908) accidentally found that a uranium-rich mineral called pitchblende emits invisible, penetrating rays that can darken a photographic plate enclosed in an opaque envelope. The rays therefore carry energy; but amazingly, the pitchblende emits them continuously without any energy input. This is an apparent violation of the law of conservation of energy, one that we now understand is due to the conversion of a small amount of mass into energy, as related in Einstein's famous equation $E = mc^2$. It was soon evident that Becquerel's rays originate in the nuclei of the atoms and have other unique characteristics. The emission of these rays is called **nuclear radioactivity** or simply **radioactivity**. The rays themselves are called **nuclear radiation**. A nucleus that spontaneously destroys part of its mass to emit radiation is said to **decay** (a term also used to describe the emission of radiation by atoms in excited states). A substance or object that emits nuclear radiation is said to be **radioactive**.

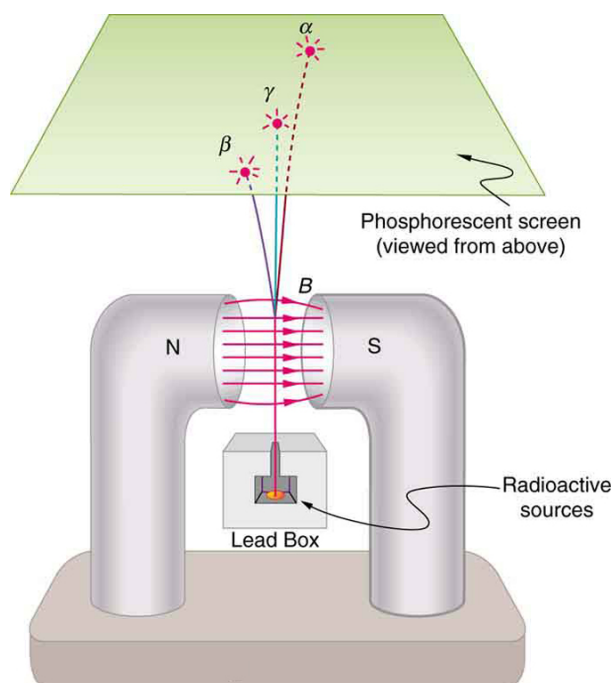
Two types of experimental evidence imply that Becquerel's rays originate deep in the heart (or nucleus) of an atom. First, the radiation is found to be associated with certain elements, such as uranium. Radiation does not vary with chemical state—that is, uranium is radioactive whether it is in the form of an element or compound. In addition, radiation does not vary with temperature, pressure, or ionization state of the uranium atom. Since all of these factors affect electrons in an atom, the radiation cannot come from electron transitions, as atomic spectra do. The huge energy emitted during each event is the second piece of evidence that the radiation cannot be atomic. Nuclear radiation has energies of the order of 10^6 eV per event, which is much greater than the typical atomic energies (a few eV), such as that observed in spectra and chemical reactions, and more than ten times as high as the most energetic characteristic x rays. Becquerel did not vigorously pursue his discovery for very long. In 1898, Marie Curie (1867–

1934), then a graduate student married the already well-known French physicist Pierre Curie (1859–1906), began her doctoral study of Becquerel's rays. She and her husband soon discovered two new radioactive elements, which she named *polonium* (after her native land) and *radium* (because it radiates). These two new elements filled holes in the periodic table and, further, displayed much higher levels of radioactivity per gram of material than uranium. Over a period of four years, working under poor conditions and spending their own funds, the Curies processed more than a ton of uranium ore to isolate a gram of radium salt. Radium became highly sought after, because it was about two million times as radioactive as uranium. Curie's radium salt glowed visibly from the radiation that took its toll on them and other unaware researchers. Shortly after completing her Ph.D., both Curies and Becquerel shared the 1903 Nobel Prize in physics for their work on radioactivity. Pierre was killed in a horse cart accident in 1906, but Marie continued her study of radioactivity for nearly 30 more years. Awarded the 1911 Nobel Prize in chemistry for her discovery of two new elements, she remains the only person to win Nobel Prizes in physics and chemistry. Marie's radioactive fingerprints on some pages of her notebooks can still expose film, and she suffered from radiation-induced lesions. She died of leukemia likely caused by radiation, but she was active in research almost until her death in 1934. The following year, her daughter and son-in-law, Irene and Frederic Joliot-Curie, were awarded the Nobel Prize in chemistry for their discovery of artificially induced radiation, adding to a remarkable family legacy.

Alpha, Beta, and Gamma

Research begun by people such as New Zealander Ernest Rutherford soon after the discovery of nuclear radiation indicated that different types of rays are emitted. Eventually, three types were distinguished and named **alpha** (α), **beta** (β), and **gamma** (γ), because, like x-rays, their identities were initially unknown. [\[link\]](#) shows what happens if the rays are passed through a magnetic field. The γ s are unaffected, while the α s and β s are deflected in opposite directions, indicating the α s are positive, the β s negative, and the γ s uncharged. Rutherford used both magnetic and electric fields to show that α s have a positive charge twice the magnitude of an electron, or $+2 |q_e|$. In the process, he found the α s charge to mass ratio to be several

thousand times smaller than the electron's. Later on, Rutherford collected α s from a radioactive source and passed an electric discharge through them, obtaining the spectrum of recently discovered helium gas. Among many important discoveries made by Rutherford and his collaborators was the proof that *α radiation is the emission of a helium nucleus*. Rutherford won the Nobel Prize in chemistry in 1908 for his early work. He continued to make important contributions until his death in 1934.



Alpha, beta, and gamma rays are passed through a magnetic field on the way to a phosphorescent screen. The α s and β s bend in opposite directions, while the γ s are unaffected, indicating a positive charge for α s, negative for β s, and neutral for γ s. Consistent results are obtained with electric fields. Collection of the radiation offers further

confirmation from the direct measurement of excess charge.

Other researchers had already proved that β s are negative and have the same mass and same charge-to-mass ratio as the recently discovered electron. By 1902, it was recognized that *β radiation is the emission of an electron*. Although β s are electrons, they do not exist in the nucleus before it decays and are not ejected atomic electrons—the electron is created in the nucleus at the instant of decay.

Since γ s remain unaffected by electric and magnetic fields, it is natural to think they might be photons. Evidence for this grew, but it was not until 1914 that this was proved by Rutherford and collaborators. By scattering γ radiation from a crystal and observing interference, they demonstrated that *γ radiation is the emission of a high-energy photon by a nucleus*. In fact, γ radiation comes from the de-excitation of a nucleus, just as an x ray comes from the de-excitation of an atom. The names " γ ray" and "x ray" identify the source of the radiation. At the same energy, γ rays and x rays are otherwise identical.

Type of Radiation	Range
α -Particles	A sheet of paper, a few cm of air, fractions of a mm of tissue

Type of Radiation	Range
β -Particles	A thin aluminum plate, or tens of cm of tissue
γ Rays	Several cm of lead or meters of concrete

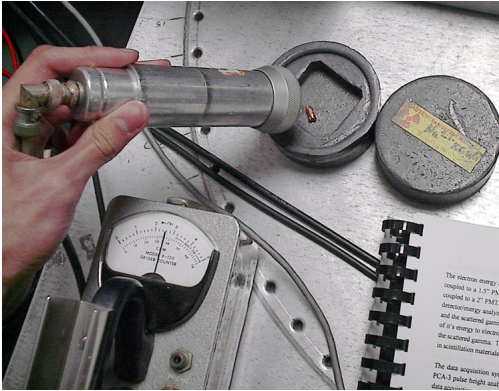
Properties of Nuclear Radiation

Ionization and Range

Two of the most important characteristics of α , β , and γ rays were recognized very early. All three types of nuclear radiation produce *ionization* in materials, but they penetrate different distances in materials—that is, they have different *ranges*. Let us examine why they have these characteristics and what are some of the consequences.

Like x rays, nuclear radiation in the form of α s, β s, and γ s has enough energy per event to ionize atoms and molecules in any material. The energy emitted in various nuclear decays ranges from a few keV to more than 10 MeV, while only a few eV are needed to produce ionization. The effects of x rays and nuclear radiation on biological tissues and other materials, such as solid state electronics, are directly related to the ionization they produce. All of them, for example, can damage electronics or kill cancer cells. In addition, methods for detecting x rays and nuclear radiation are based on ionization, directly or indirectly. All of them can ionize the air between the plates of a capacitor, for example, causing it to discharge. This is the basis of inexpensive personal radiation monitors, such as pictured in [\[link\]](#). Apart from α , β , and γ , there are other forms of nuclear radiation as well, and these also produce ionization with similar effects. We define **ionizing radiation** as any form of radiation that produces ionization

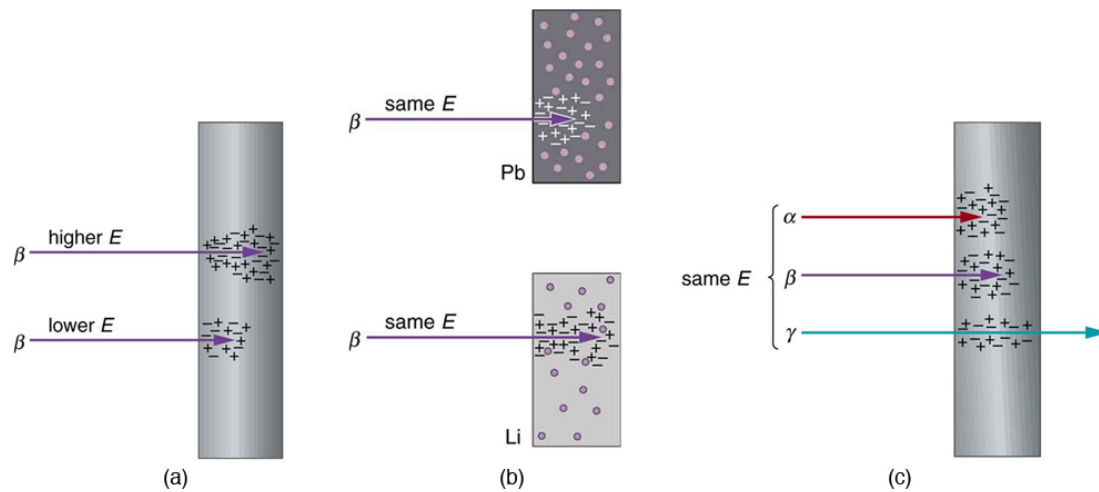
whether nuclear in origin or not, since the effects and detection of the radiation are related to ionization.



These dosimeters (literally, dose meters) are personal radiation monitors that detect the amount of radiation by the discharge of a rechargeable internal capacitor. The amount of discharge is related to the amount of ionizing radiation encountered, a measurement of dose. One dosimeter is shown in the charger. Its scale is read through an eyepiece on the top. (credit: L. Chang, Wikimedia Commons)

The **range of radiation** is defined to be the distance it can travel through a material. Range is related to several factors, including the energy of the

radiation, the material encountered, and the type of radiation (see [\[link\]](#)). The higher the *energy*, the greater the range, all other factors being the same. This makes good sense, since radiation loses its energy in materials primarily by producing ionization in them, and each ionization of an atom or a molecule requires energy that is removed from the radiation. The amount of ionization is, thus, directly proportional to the energy of the particle of radiation, as is its range.



The penetration or range of radiation depends on its energy, the material it encounters, and the type of radiation. (a) Greater energy means greater range. (b) Radiation has a smaller range in materials with high electron density. (c) Alphas have the smallest range, betas have a greater range, and gammas penetrate the farthest.

Radiation can be absorbed or shielded by materials, such as the lead aprons dentists drape on us when taking x rays. Lead is a particularly effective shield compared with other materials, such as plastic or air. How does the range of radiation depend on *material*? Ionizing radiation interacts best with charged particles in a material. Since electrons have small masses, they most readily absorb the energy of the radiation in collisions. The greater the

density of a material and, in particular, the greater the density of electrons within a material, the smaller the range of radiation.

Note:

Collisions

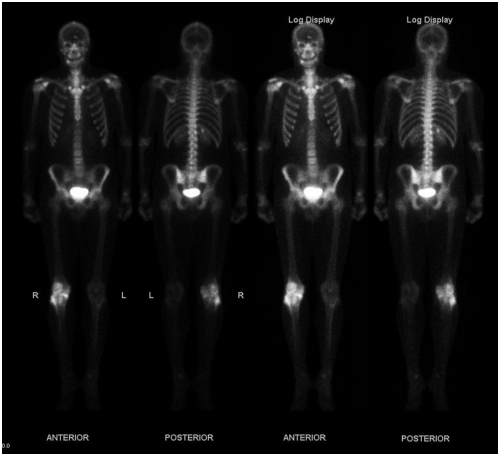
Conservation of energy and momentum often results in energy transfer to a less massive object in a collision. This was discussed in detail in [Work, Energy, and Energy Resources](#), for example.

Different *types* of radiation have different ranges when compared at the same energy and in the same material. Alphas have the shortest range, betas penetrate farther, and gammas have the greatest range. This is directly related to charge and speed of the particle or type of radiation. At a given energy, each α , β , or γ will produce the same number of ionizations in a material (each ionization requires a certain amount of energy on average). The more readily the particle produces ionization, the more quickly it will lose its energy. The effect of *charge* is as follows: The α has a charge of $+2q_e$, the β has a charge of $-q_e$, and the γ is uncharged. The electromagnetic force exerted by the α is thus twice as strong as that exerted by the β and it is more likely to produce ionization. Although chargeless, the γ does interact weakly because it is an electromagnetic wave, but it is less likely to produce ionization in any encounter. More quantitatively, the change in momentum Δp given to a particle in the material is $\Delta p = F\Delta t$, where F is the force the α , β , or γ exerts over a time Δt . The smaller the charge, the smaller is F and the smaller is the momentum (and energy) lost. Since the speed of alphas is about 5% to 10% of the speed of light, classical (non-relativistic) formulas apply.

The *speed* at which they travel is the other major factor affecting the range of α s, β s, and γ s. The faster they move, the less time they spend in the vicinity of an atom or a molecule, and the less likely they are to interact. Since α s and β s are particles with mass (helium nuclei and electrons, respectively), their energy is kinetic, given classically by $\frac{1}{2}mv^2$. The mass

of the β particle is thousands of times less than that of the α s, so that β s must travel much faster than α s to have the same energy. Since β s move faster (most at relativistic speeds), they have less time to interact than α s. Gamma rays are photons, which must travel at the speed of light. They are even less likely to interact than a β , since they spend even less time near a given atom (and they have no charge). The range of γ s is thus greater than the range of β s.

Alpha radiation from radioactive sources has a range much less than a millimeter of biological tissues, usually not enough to even penetrate the dead layers of our skin. On the other hand, the same α radiation can penetrate a few centimeters of air, so mere distance from a source prevents α radiation from reaching us. This makes α radiation relatively safe for our body compared to β and γ radiation. Typical β radiation can penetrate a few millimeters of tissue or about a meter of air. Beta radiation is thus hazardous even when not ingested. The range of β s in lead is about a millimeter, and so it is easy to store β sources in lead radiation-proof containers. Gamma rays have a much greater range than either α s or β s. In fact, if a given thickness of material, like a lead brick, absorbs 90% of the γ s, then a second lead brick will only absorb 90% of what got through the first. Thus, γ s do not have a well-defined range; we can only cut down the amount that gets through. Typically, γ s can penetrate many meters of air, go right through our bodies, and are effectively shielded (that is, reduced in intensity to acceptable levels) by many centimeters of lead. One benefit of γ s is that they can be used as radioactive tracers (see [\[link\]](#)).



This image of the concentration of a radioactive tracer in a patient's body reveals where the most active bone cells are, an indication of bone cancer. A short-lived radioactive substance that locates itself selectively is given to the patient, and the radiation is measured with an external detector. The emitted γ radiation has a sufficient range to leave the body—the range of α s and β s is too small for them to be observed outside the patient. (credit: Kieran Maher, Wikimedia Commons)

Note:**PhET Explorations: Beta Decay**

Build an atom out of protons, neutrons, and electrons, and see how the element, charge, and mass change. Then play a game to test your ideas!

<https://archive.cnx.org/specials/f0a27b96-f5c8-11e5-a22c-73f8c149bebf/beta-decay/#sim-multiple-atoms>

Section Summary

- Some nuclei are radioactive—they spontaneously decay destroying some part of their mass and emitting energetic rays, a process called nuclear radioactivity.
- Nuclear radiation, like x rays, is ionizing radiation, because energy sufficient to ionize matter is emitted in each decay.
- The range (or distance traveled in a material) of ionizing radiation is directly related to the charge of the emitted particle and its energy, with greater-charge and lower-energy particles having the shortest ranges.
- Radiation detectors are based directly or indirectly upon the ionization created by radiation, as are the effects of radiation on living and inert materials.

Conceptual Questions

Exercise:**Problem:**

Suppose the range for 5.0 MeV α ray is known to be 2.0 mm in a certain material. Does this mean that every 5.0 MeV α ray that strikes this material travels 2.0 mm, or does the range have an average value with some statistical fluctuations in the distances traveled? Explain.

Exercise:

Problem:

What is the difference between γ rays and characteristic x rays? Is either necessarily more energetic than the other? Which can be the most energetic?

Exercise:**Problem:**

Ionizing radiation interacts with matter by scattering from electrons and nuclei in the substance. Based on the law of conservation of momentum and energy, explain why electrons tend to absorb more energy than nuclei in these interactions.

Exercise:**Problem:**

What characteristics of radioactivity show it to be nuclear in origin and not atomic?

Exercise:**Problem:**

What is the source of the energy emitted in radioactive decay? Identify an earlier conservation law, and describe how it was modified to take such processes into account.

Exercise:**Problem:**

Consider [\[link\]](#). If an electric field is substituted for the magnetic field with positive charge instead of the north pole and negative charge instead of the south pole, in which directions will the α , β , and γ rays bend?

Exercise:

Problem:

Explain how an α particle can have a larger range in air than a β particle with the same energy in lead.

Exercise:**Problem:**

Arrange the following according to their ability to act as radiation shields, with the best first and worst last. Explain your ordering in terms of how radiation loses its energy in matter.

- (a) A solid material with low density composed of low-mass atoms.
- (b) A gas composed of high-mass atoms.
- (c) A gas composed of low-mass atoms.
- (d) A solid with high density composed of high-mass atoms.

Exercise:**Problem:**

Often, when people have to work around radioactive materials spills, we see them wearing white coveralls (usually a plastic material). What types of radiation (if any) do you think these suits protect the worker from, and how?

Glossary

alpha rays

one of the types of rays emitted from the nucleus of an atom

beta rays

one of the types of rays emitted from the nucleus of an atom

gamma rays

one of the types of rays emitted from the nucleus of an atom

ionizing radiation

radiation (whether nuclear in origin or not) that produces ionization
whether nuclear in origin or not

nuclear radiation

rays that originate in the nuclei of atoms, the first examples of which
were discovered by Becquerel

radioactivity

the emission of rays from the nuclei of atoms

radioactive

a substance or object that emits nuclear radiation

range of radiation

the distance that the radiation can travel through a material

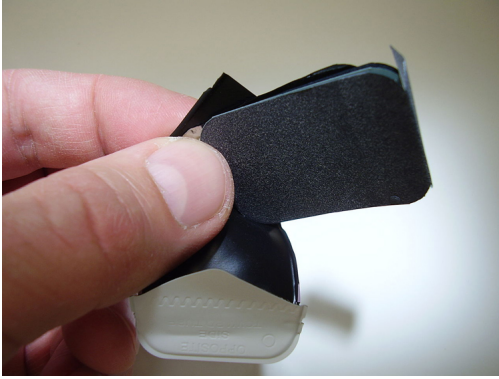
Radiation Detection and Detectors

- Explain the working principle of a Geiger tube.
- Define and discuss radiation detectors.

It is well known that ionizing radiation affects us but does not trigger nerve impulses. Newspapers carry stories about unsuspecting victims of radiation poisoning who fall ill with radiation sickness, such as burns and blood count changes, but who never felt the radiation directly. This makes the detection of radiation by instruments more than an important research tool. This section is a brief overview of radiation detection and some of its applications.

Human Application

The first direct detection of radiation was Becquerel's fogged photographic plate. Photographic film is still the most common detector of ionizing radiation, being used routinely in medical and dental x rays. Nuclear radiation is also captured on film, such as seen in [\[link\]](#). The mechanism for film exposure by ionizing radiation is similar to that by photons. A quantum of energy interacts with the emulsion and alters it chemically, thus exposing the film. The quantum come from an α -particle, β -particle, or photon, provided it has more than the few eV of energy needed to induce the chemical change (as does all ionizing radiation). The process is not 100% efficient, since not all incident radiation interacts and not all interactions produce the chemical change. The amount of film darkening is related to exposure, but the darkening also depends on the type of radiation, so that absorbers and other devices must be used to obtain energy, charge, and particle-identification information.

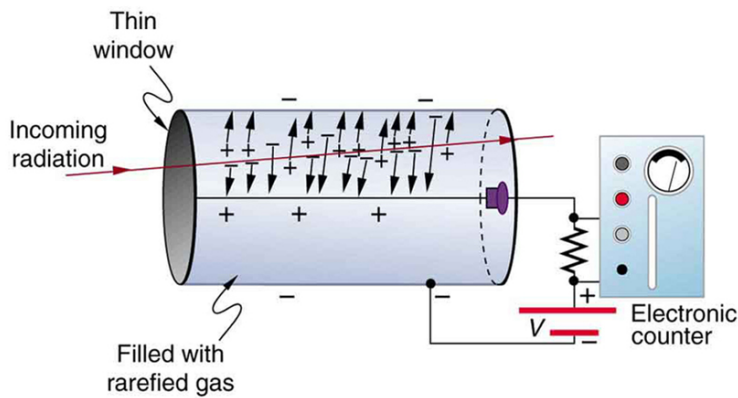


Film badges contain film similar to that used in this dental x-ray film and is sandwiched between various absorbers to determine the penetrating ability of the radiation as well as the amount.
(credit: Werneuchen, Wikimedia Commons)

Another very common **radiation detector** is the **Geiger tube**. The clicking and buzzing sound we hear in dramatizations and documentaries, as well as in our own physics labs, is usually an audio output of events detected by a Geiger counter. These relatively inexpensive radiation detectors are based on the simple and sturdy Geiger tube, shown schematically in [\[link\]](#)(b). A conducting cylinder with a wire along its axis is filled with an insulating gas so that a voltage applied between the cylinder and wire produces almost no current. Ionizing radiation passing through the tube produces free ion pairs that are attracted to the wire and cylinder, forming a current that is detected as a count. The word count implies that there is no information on energy, charge, or type of radiation with a simple Geiger counter. They do not detect every particle, since some radiation can pass through without producing enough ionization to be detected. However, Geiger counters are very useful in producing a prompt output that reveals the existence and relative intensity of ionizing radiation.



(a)

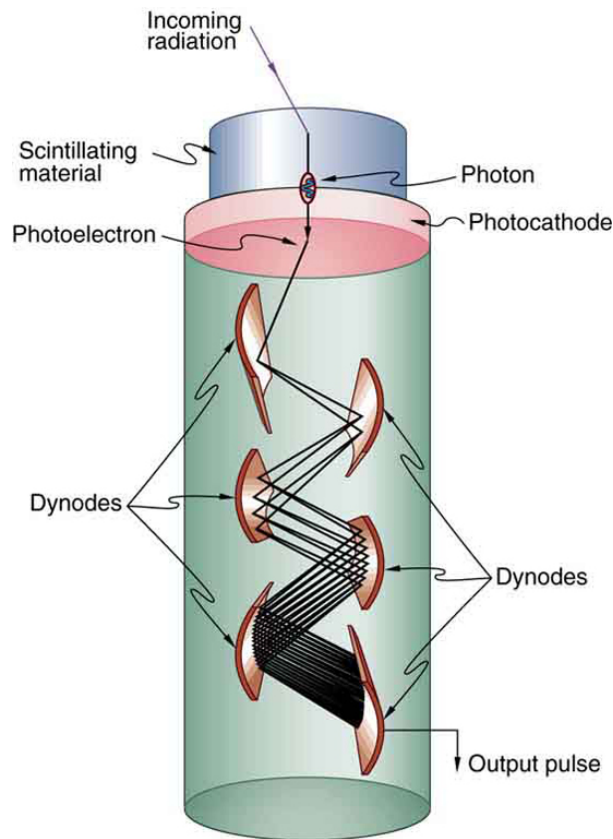


(b)

(a) Geiger counters such as this one are used for prompt monitoring of radiation levels, generally giving only relative intensity and not identifying the type or energy of the radiation. (credit: TimVickers, Wikimedia Commons) (b) Voltage applied between the cylinder and wire in a Geiger tube causes ions and electrons produced by radiation passing through the gas-filled cylinder to move towards them. The resulting current is detected and registered as a count.

Another radiation detection method records light produced when radiation interacts with materials. The energy of the radiation is sufficient to excite atoms in a material that may fluoresce, such as the phosphor used by Rutherford's group. Materials called **scintillators** use a more complex collaborative process to convert radiation energy into light. Scintillators may be liquid or solid, and they can be very efficient. Their light output can provide information about the energy, charge, and type of radiation. Scintillator light flashes are very brief in duration, enabling the detection of a huge number of particles in short periods of time. Scintillator detectors are used in a variety of research and diagnostic applications. Among these are the detection by satellite-mounted equipment of the radiation from distant galaxies, the analysis of radiation from a person indicating body burdens, and the detection of exotic particles in accelerator laboratories.

Light from a scintillator is converted into electrical signals by devices such as the **photomultiplier** tube shown schematically in [\[link\]](#). These tubes are based on the photoelectric effect, which is multiplied in stages into a cascade of electrons, hence the name photomultiplier. Light entering the photomultiplier strikes a metal plate, ejecting an electron that is attracted by a positive potential difference to the next plate, giving it enough energy to eject two or more electrons, and so on. The final output current can be made proportional to the energy of the light entering the tube, which is in turn proportional to the energy deposited in the scintillator. Very sophisticated information can be obtained with scintillators, including energy, charge, particle identification, direction of motion, and so on.



Photomultipliers use the photoelectric effect on the photocathode to convert the light output of a scintillator into an electrical signal. Each successive dynode has a more-positive potential than the last and attracts the ejected electrons, giving them more energy. The number of electrons is thus multiplied at each dynode, resulting in an easily detected output current.

Solid-state radiation detectors convert ionization produced in a semiconductor (like those found in computer chips) directly into an

electrical signal. Semiconductors can be constructed that do not conduct current in one particular direction. When a voltage is applied in that direction, current flows only when ionization is produced by radiation, similar to what happens in a Geiger tube. Further, the amount of current in a solid-state detector is closely related to the energy deposited and, since the detector is solid, it can have a high efficiency (since ionizing radiation is stopped in a shorter distance in solids fewer particles escape detection). As with scintillators, very sophisticated information can be obtained from solid-state detectors.

Note:

PhET Explorations: Radioactive Dating Game

Learn about different types of radiometric dating, such as carbon dating. Understand how decay and half life work to enable radiometric dating to work. Play a game that tests your ability to match the percentage of the dating element that remains to the age of the object.

<https://archive.cnx.org/specials/d709a8b0-068c-11e6-bcfb-f38266817c66/radioactive-dating-game/#sim-half-life>

Section Summary

- Radiation detectors are based directly or indirectly upon the ionization created by radiation, as are the effects of radiation on living and inert materials.

Conceptual Questions

Exercise:

Problem:

Is it possible for light emitted by a scintillator to be too low in frequency to be used in a photomultiplier tube? Explain.

Problems & Exercises

Exercise:

Problem:

The energy of 30.0 eV is required to ionize a molecule of the gas inside a Geiger tube, thereby producing an ion pair. Suppose a particle of ionizing radiation deposits 0.500 MeV of energy in this Geiger tube. What maximum number of ion pairs can it create?

Solution:

$$1.67 \times 10^4$$

Exercise:

Problem:

A particle of ionizing radiation creates 4000 ion pairs in the gas inside a Geiger tube as it passes through. What minimum energy was deposited, if 30.0 eV is required to create each ion pair?

Exercise:

Problem:

(a) Repeat [\[link\]](#), and convert the energy to joules or calories. (b) If all of this energy is converted to thermal energy in the gas, what is its temperature increase, assuming 50.0 cm³ of ideal gas at 0.250-atm pressure? (The small answer is consistent with the fact that the energy is large on a quantum mechanical scale but small on a macroscopic scale.)

Exercise:

Problem:

Suppose a particle of ionizing radiation deposits 1.0 MeV in the gas of a Geiger tube, all of which goes to creating ion pairs. Each ion pair requires 30.0 eV of energy. (a) The applied voltage sweeps the ions out of the gas in $1.00\ \mu\text{s}$. What is the current? (b) This current is smaller than the actual current since the applied voltage in the Geiger tube accelerates the separated ions, which then create other ion pairs in subsequent collisions. What is the current if this last effect multiplies the number of ion pairs by 900?

Glossary**Geiger tube**

a very common radiation detector that usually gives an audio output

photomultiplier

a device that converts light into electrical signals

radiation detector

a device that is used to detect and track the radiation from a radioactive reaction

scintillators

a radiation detection method that records light produced when radiation interacts with materials

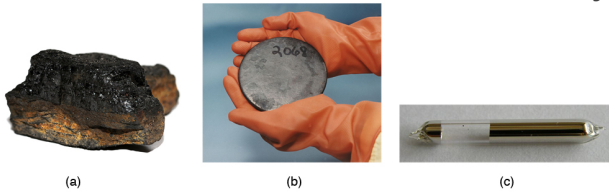
solid-state radiation detectors

semiconductors fabricated to directly convert incident radiation into electrical current

Substructure of the Nucleus

- Define and discuss the nucleus in an atom.
- Define atomic number.
- Define and discuss isotopes.
- Calculate the density of the nucleus.
- Explain nuclear force.

What is inside the nucleus? Why are some nuclei stable while others decay? (See [\[link\]](#).) Why are there different types of decay (α , β and γ)? Why are nuclear decay energies so large? Pursuing natural questions like these has led to far more fundamental discoveries than you might imagine.



Why is most of the carbon in this coal stable (a), while the uranium in the disk (b) slowly decays over billions of years? Why is cesium in this ampule (c) even less stable than the uranium, decaying in far less than 1/1,000,000 the time? What is the reason uranium and cesium undergo different types of decay (α and β , respectively)?

(credits: (a) Bresson Thomas, Wikimedia Commons; (b) U.S. Department of Energy; (c) Tomihahndorf, Wikimedia Commons)

We have already identified **protons** as the particles that carry positive charge in the nuclei. However, there are actually *two* types of particles in the nuclei—the *proton* and the *neutron*, referred to collectively as **nucleons**, the constituents of nuclei. As its name implies, the **neutron** is a neutral particle ($q = 0$) that has

nearly the same mass and intrinsic spin as the proton. [\[link\]](#) compares the masses of protons, neutrons, and electrons. Note how close the proton and neutron masses are, but the neutron is slightly more massive once you look past the third digit. Both nucleons are much more massive than an electron. In fact, $m_p = 1836m_e$ (as noted in [Medical Applications of Nuclear Physics](#) and $m_n = 1839m_e$.

[\[link\]](#) also gives masses in terms of mass units that are more convenient than kilograms on the atomic and nuclear scale. The first of these is the *unified atomic mass unit* (u), defined as

Equation:

$$1 \text{ u} = 1.6605 \times 10^{-27} \text{ kg}.$$

This unit is defined so that a neutral carbon ^{12}C atom has a mass of exactly 12 u. Masses are also expressed in units of MeV/c^2 . These units are very convenient when considering the conversion of mass into energy (and vice versa), as is so prominent in nuclear processes. Using $E = mc^2$ and units of m in MeV/c^2 , we find that c^2 cancels and E comes out conveniently in MeV. For example, if the rest mass of a proton is converted entirely into energy, then

Equation:

$$E = mc^2 = (938.27 \text{ MeV}/c^2)c^2 = 938.27 \text{ MeV}.$$

It is useful to note that 1 u of mass converted to energy produces 931.5 MeV, or

Equation:

$$1 \text{ u} = 931.5 \text{ MeV}/c^2.$$

All properties of a nucleus are determined by the number of protons and neutrons it has. A specific combination of protons and neutrons is called a **nuclide** and is a unique nucleus. The following notation is used to represent a particular nuclide:

Equation:

$${}^A_Z\text{X}_N,$$

where the symbols A , X , Z , and N are defined as follows: The *number of protons in a nucleus* is the **atomic number** Z , as defined in [Medical Applications of Nuclear Physics](#). X is the *symbol for the element*, such as Ca for calcium. However, once Z is known, the element is known; hence, Z and X are redundant. For example, $Z = 20$ is always calcium, and calcium always has $Z = 20$. N is the *number of neutrons* in a nucleus. In the notation for a nuclide, the subscript N is usually omitted. The symbol A is defined as the number of nucleons or the *total number of protons and neutrons*,

Equation:

$$A = N + Z,$$

where A is also called the **mass number**. This name for A is logical; the mass of an atom is nearly equal to the mass of its nucleus, since electrons have so little mass. The mass of the nucleus turns out to be nearly equal to the sum of the masses of the protons and neutrons in it, which is proportional to A . In this context, it is particularly convenient to express masses in units of u. Both protons and neutrons have masses close to 1 u, and so the mass of an atom is close to A u. For example, in an oxygen nucleus with eight protons and eight neutrons, $A = 16$, and its mass is 16 u. As noticed, the unified atomic mass unit is defined so that a neutral carbon atom (actually a ^{12}C atom) has a mass of *exactly* 12 u. Carbon was chosen as the standard, partly because of its importance in organic chemistry (see [Appendix A](#)).

Particle	Symbol	kg	u	MeV c^2
Proton	p	1.67262×10^{-27}	1.007276	938.27
Neutron	n	1.67493×10^{-27}	1.008665	939.57

Particle	Symbol	kg	u	MeVc ²
Electron	e	9.1094×10^{-31}	0.00054858	0.511

Masses of the Proton, Neutron, and Electron

Let us look at a few examples of nuclides expressed in the ${}^A_Z\text{X}_N$ notation. The nucleus of the simplest atom, hydrogen, is a single proton, or ${}^1_1\text{H}$ (the zero for no neutrons is often omitted). To check this symbol, refer to the periodic table—you see that the atomic number Z of hydrogen is 1. Since you are given that there are no neutrons, the mass number A is also 1. Suppose you are told that the helium nucleus or α particle has two protons and two neutrons. You can then see that it is written ${}^4_2\text{He}_2$. There is a scarce form of hydrogen found in nature called deuterium; its nucleus has one proton and one neutron and, hence, twice the mass of common hydrogen. The symbol for deuterium is, thus, ${}^2_1\text{H}_1$ (sometimes D is used, as for deuterated water D_2O). An even rarer—and radioactive—form of hydrogen is called tritium, since it has a single proton and two neutrons, and it is written ${}^3_1\text{H}_2$. These three varieties of hydrogen have nearly identical chemistries, but the nuclei differ greatly in mass, stability, and other characteristics. Nuclei (such as those of hydrogen) having the same Z and different N s are defined to be **isotopes** of the same element.

There is some redundancy in the symbols A , X , Z , and N . If the element X is known, then Z can be found in a periodic table and is always the same for a given element. If both A and X are known, then N can also be determined (first find Z ; then, $N = A - Z$). Thus the simpler notation for nuclides is

Equation:

$${}^A\text{X},$$

which is sufficient and is most commonly used. For example, in this simpler notation, the three isotopes of hydrogen are ${}^1\text{H}$, ${}^2\text{H}$, and ${}^3\text{H}$, while the α particle is ${}^4\text{He}$. We read this backward, saying helium-4 for ${}^4\text{He}$, or uranium-238 for ${}^{238}\text{U}$. So for ${}^{238}\text{U}$, should we need to know, we can determine that $Z = 92$ for uranium from the periodic table, and, thus, $N = 238 - 92 = 146$.

A variety of experiments indicate that a nucleus behaves something like a tightly packed ball of nucleons, as illustrated in [\[link\]](#). These nucleons have large kinetic energies and, thus, move rapidly in very close contact. Nucleons can be separated by a large force, such as in a collision with another nucleus, but resist strongly being pushed closer together. The most compelling evidence that nucleons are closely packed in a nucleus is that the **radius of a nucleus**, r , is found to be given approximately by

Equation:

$$r = r_0 A^{1/3},$$

where $r_0 = 1.2$ fm and A is the mass number of the nucleus. Note that $r^3 \propto A$. Since many nuclei are spherical, and the volume of a sphere is $V = (4/3)\pi r^3$, we see that $V \propto A$ —that is, the volume of a nucleus is proportional to the number of nucleons in it. This is what would happen if you pack nucleons so closely that there is no empty space between them.



A model of the
nucleus.

Nucleons are held together by nuclear forces and resist both being pulled apart and pushed inside one another. The volume of the nucleus is the sum of the volumes of the nucleons in it, here shown in different colors to represent protons and neutrons.

Example:

How Small and Dense Is a Nucleus?

(a) Find the radius of an iron-56 nucleus. (b) Find its approximate density in kg/m^3 , approximating the mass of ^{56}Fe to be 56 u.

Strategy and Concept

(a) Finding the radius of ^{56}Fe is a straightforward application of $r = r_0 A^{1/3}$, given $A = 56$. (b) To find the approximate density, we assume the nucleus is spherical (this one actually is), calculate its volume using the radius found in part (a), and then find its density from $\rho = m / V$. Finally, we will need to convert density from units of u/fm^3 to kg/m^3 .

Solution

(a) The radius of a nucleus is given by

Equation:

$$r = r_0 A^{1/3}.$$

Substituting the values for r_0 and A yields

Equation:

$$\begin{aligned} r &= (1.2 \text{ fm})(56)^{1/3} = (1.2 \text{ fm})(3.83) \\ &= 4.6 \text{ fm}. \end{aligned}$$

(b) Density is defined to be $\rho = m / V$, which for a sphere of radius r is

Equation:

$$\rho = \frac{m}{V} = \frac{m}{(4/3)\pi r^3}.$$

Substituting known values gives

Equation:

$$\begin{aligned} \rho &= \frac{56 \text{ u}}{(1.33)(3.14)(4.6 \text{ fm})^3} \\ &= 0.138 \text{ u}/\text{fm}^3. \end{aligned}$$

Converting to units of kg/m^3 , we find

Equation:

$$\begin{aligned} \rho &= (0.138 \text{ u}/\text{fm}^3)(1.66 \times 10^{-27} \text{ kg}/\text{u})\left(\frac{1 \text{ fm}}{10^{-15} \text{ m}}\right) \\ &= 2.3 \times 10^{17} \text{ kg}/\text{m}^3. \end{aligned}$$

Discussion

(a) The radius of this medium-sized nucleus is found to be approximately 4.6 fm, and so its diameter is about 10 fm, or 10^{-14} m. In our discussion of Rutherford's discovery of the nucleus, we noticed that it is about 10^{-15} m in diameter (which is for lighter nuclei), consistent with this result to an order of magnitude. The nucleus is much smaller in diameter than the typical atom, which has a diameter of the order of 10^{-10} m.

(b) The density found here is so large as to cause disbelief. It is consistent with earlier discussions we have had about the nucleus being very small and containing nearly all of the mass of the atom. Nuclear densities, such as found here, are about 2×10^{14} times greater than that of water, which has a density of “only” 10^3 kg/m³. One cubic meter of nuclear matter, such as found in a neutron star, has the same mass as a cube of water 61 km on a side.

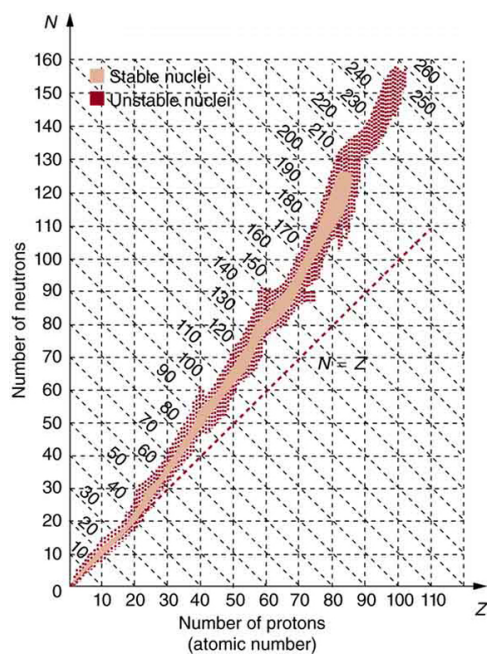
Nuclear Forces and Stability

What forces hold a nucleus together? The nucleus is very small and its protons, being positive, exert tremendous repulsive forces on one another. (The Coulomb force increases as charges get closer, since it is proportional to $1/r^2$, even at the tiny distances found in nuclei.) The answer is that two previously unknown forces hold the nucleus together and make it into a tightly packed ball of nucleons. These forces are called the *weak and strong nuclear forces*.

Nuclear forces are so short ranged that they fall to zero strength when nucleons are separated by only a few fm. However, like glue, they are strongly attracted when the nucleons get close to one another. The strong nuclear force is about 100 times more attractive than the repulsive EM force, easily holding the nucleons together. Nuclear forces become extremely repulsive if the nucleons get too close, making nucleons strongly resist being pushed inside one another, something like ball bearings.

The fact that nuclear forces are very strong is responsible for the very large energies emitted in nuclear decay. During decay, the forces do work, and since work is force times the distance ($W = Fd \cos \theta$), a large force can result in a large emitted energy. In fact, we know that there are *two* distinct nuclear forces because of the different types of nuclear decay—the strong nuclear force is responsible for α decay, while the weak nuclear force is responsible for β decay.

The many stable and unstable nuclei we have explored, and the hundreds we have not discussed, can be arranged in a table called the **chart of the nuclides**, a simplified version of which is shown in [\[link\]](#). Nuclides are located on a plot of N versus Z . Examination of a detailed chart of the nuclides reveals patterns in the characteristics of nuclei, such as stability, abundance, and types of decay, analogous to but more complex than the systematics in the periodic table of the elements.



Simplified chart of the nuclides, a graph of N versus Z for known nuclides. The patterns of stable and unstable nuclides reveal characteristics of the nuclear forces. The dashed line is for $N = Z$. Numbers along diagonals are mass numbers A .

In principle, a nucleus can have any combination of protons and neutrons, but [\[link\]](#) shows a definite pattern for those that are stable. For low-mass nuclei, there is a strong tendency for N and Z to be nearly equal. This means that the nuclear force is more attractive when $N = Z$. More detailed examination reveals greater stability when N and Z are even numbers—nuclear forces are more attractive when neutrons and protons are in pairs. For increasingly higher masses, there are progressively more neutrons than protons in stable nuclei. This is due to the ever-growing repulsion between protons. Since nuclear forces are short ranged, and the Coulomb force is long ranged, an excess of neutrons keeps the protons a little farther apart, reducing Coulomb repulsion. Decay modes of nuclides out of the region of stability consistently produce nuclides closer to the region of stability. There are more stable nuclei having certain numbers of protons and neutrons, called **magic numbers**. Magic numbers indicate a shell structure for the nucleus in which closed shells are more stable. Nuclear shell theory has been very successful in explaining nuclear energy levels, nuclear decay, and the greater stability of nuclei with closed shells. We have been producing ever-heavier transuranic elements since the early 1940s, and we have now produced the element with $Z = 118$. There are theoretical predictions of an island of relative stability for nuclei with such high Z s.



The German-born
American
physicist Maria
Goeppert Mayer
(1906–1972)

shared the 1963 Nobel Prize in physics with J. Jensen for the creation of the nuclear shell model. This successful nuclear model has nucleons filling shells analogous to electron shells in atoms. It was inspired by patterns observed in nuclear properties. (credit: Nobel Foundation via Wikimedia Commons)

Section Summary

- Two particles, both called nucleons, are found inside nuclei. The two types of nucleons are protons and neutrons; they are very similar, except that the proton is positively charged while the neutron is neutral. Some of their characteristics are given in [\[link\]](#) and compared with those of the electron. A mass unit convenient to atomic and nuclear processes is the unified atomic mass unit (u), defined to be

Equation:

$$1 \text{ u} = 1.6605 \times 10^{-27} \text{ kg} = 931.46 \text{ MeV}/c^2.$$

- A nuclide is a specific combination of protons and neutrons, denoted by
- Equation:**

$${}^A_Z\text{X}_N \text{ or simply } {}^A\text{X},$$

Z is the number of protons or atomic number, X is the symbol for the element, N is the number of neutrons, and A is the mass number or the total number of protons and neutrons,

Equation:

$$A = N + Z.$$

- Nuclides having the same Z but different N are isotopes of the same element.
- The radius of a nucleus, r , is approximately

Equation:

$$r = r_0 A^{1/3},$$

where $r_0 = 1.2$ fm. Nuclear volumes are proportional to A . There are two nuclear forces, the weak and the strong. Systematics in nuclear stability seen on the chart of the nuclides indicate that there are shell closures in nuclei for values of Z and N equal to the magic numbers, which correspond to highly stable nuclei.

Conceptual Questions

Exercise:

Problem:

The weak and strong nuclear forces are basic to the structure of matter. Why we do not experience them directly?

Exercise:

Problem:

Define and make clear distinctions between the terms neutron, nucleon, nucleus, nuclide, and neutrino.

Exercise:

Problem:

What are isotopes? Why do different isotopes of the same element have similar chemistries?

Problems & Exercises**Exercise:****Problem:**

Verify that a 2.3×10^{17} kg mass of water at normal density would make a cube 60 km on a side, as claimed in [\[link\]](#). (This mass at nuclear density would make a cube 1.0 m on a side.)

Solution:**Equation:**

$$\begin{aligned} m = \rho V = \rho d^3 &\Rightarrow a = \left(\frac{m}{\rho}\right)^{1/3} = \left(\frac{2.3 \times 10^{17} \text{ kg}}{1000 \text{ kg/m}^3}\right)^{\frac{1}{3}} \\ &= 61 \times 10^3 \text{ m} = 61 \text{ km} \end{aligned}$$

Exercise:**Problem:**

Find the length of a side of a cube having a mass of 1.0 kg and the density of nuclear matter, taking this to be $2.3 \times 10^{17} \text{ kg/m}^3$.

Exercise:

Problem: What is the radius of an α particle?

Solution:

1.9 fm

Exercise:

Problem:

Find the radius of a ^{238}Pu nucleus. ^{238}Pu is a manufactured nuclide that is used as a power source on some space probes.

Exercise:**Problem:**

- (a) Calculate the radius of ^{58}Ni , one of the most tightly bound stable nuclei.
- (b) What is the ratio of the radius of ^{58}Ni to that of ^{258}Ha , one of the largest nuclei ever made? Note that the radius of the largest nucleus is still much smaller than the size of an atom.
-

Solution:

- (a) 4.6 fm
- (b) 0.61 to 1

Exercise:**Problem:**

The unified atomic mass unit is defined to be $1\text{ u} = 1.6605 \times 10^{-27}\text{ kg}$. Verify that this amount of mass converted to energy yields 931.5 MeV. Note that you must use four-digit or better values for c and $|q_e|$.

Exercise:**Problem:**

What is the ratio of the velocity of a β particle to that of an α particle, if they have the same nonrelativistic kinetic energy?

Solution:

85.4 to 1

Exercise:

Problem:

If a 1.50-cm-thick piece of lead can absorb 90.0% of the γ rays from a radioactive source, how many centimeters of lead are needed to absorb all but 0.100% of the γ rays?

Exercise:**Problem:**

The detail observable using a probe is limited by its wavelength. Calculate the energy of a γ -ray photon that has a wavelength of 1×10^{-16} m, small enough to detect details about one-tenth the size of a nucleon. Note that a photon having this energy is difficult to produce and interacts poorly with the nucleus, limiting the practicability of this probe.

Solution:

12.4 GeV

Exercise:**Problem:**

(a) Show that if you assume the average nucleus is spherical with a radius $r = r_0 A^{1/3}$, and with a mass of A u, then its density is independent of A .

(b) Calculate that density in u/fm³ and kg/m³, and compare your results with those found in [\[link\]](#) for ⁵⁶Fe.

Exercise:**Problem:**

What is the ratio of the velocity of a 5.00-MeV β ray to that of an α particle with the same kinetic energy? This should confirm that β s travel much faster than α s even when relativity is taken into consideration. (See also [\[link\]](#).)

Solution:

19.3 to 1

Exercise:

Problem:

(a) What is the kinetic energy in MeV of a β ray that is traveling at $0.998c$? This gives some idea of how energetic a β ray must be to travel at nearly the same speed as a γ ray. (b) What is the velocity of the γ ray relative to the β ray?

Glossary

atomic mass

the total mass of the protons, neutrons, and electrons in a single atom

atomic number

number of protons in a nucleus

chart of the nuclides

a table comprising stable and unstable nuclei

isotopes

nuclei having the same Z and different N s

magic numbers

a number that indicates a shell structure for the nucleus in which closed shells are more stable

mass number

number of nucleons in a nucleus

neutron

a neutral particle that is found in a nucleus

nucleons

the particles found inside nuclei

nucleus

a region consisting of protons and neutrons at the center of an atom

nuclide

a type of atom whose nucleus has specific numbers of protons and neutrons

protons

the positively charged nucleons found in a nucleus

radius of a nucleus

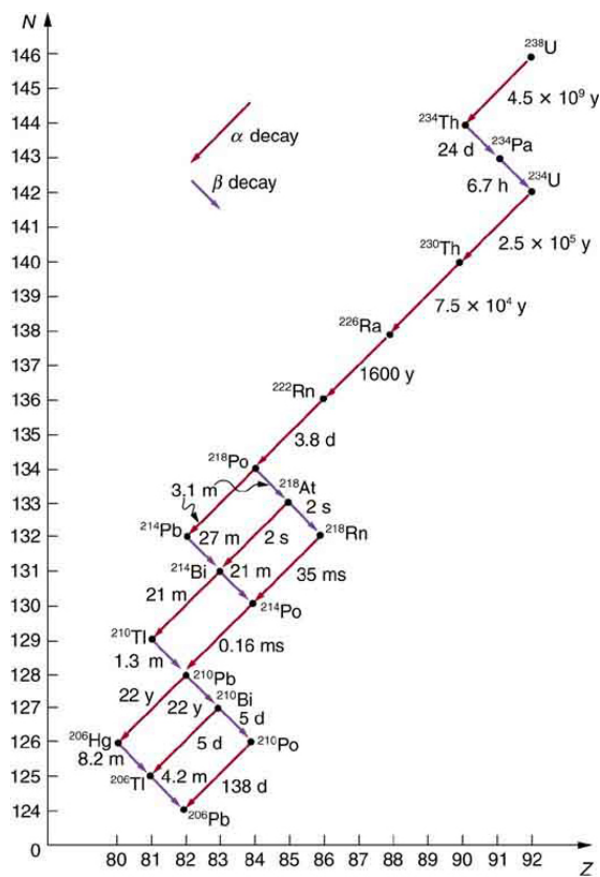
the radius of a nucleus is $r = r_0 A^{1/3}$

Nuclear Decay and Conservation Laws

- Define and discuss nuclear decay.
- State the conservation laws.
- Explain parent and daughter nucleus.
- Calculate the energy emitted during nuclear decay.

Nuclear **decay** has provided an amazing window into the realm of the very small. Nuclear decay gave the first indication of the connection between mass and energy, and it revealed the existence of two of the four basic forces in nature. In this section, we explore the major modes of nuclear decay; and, like those who first explored them, we will discover evidence of previously unknown particles and conservation laws.

Some nuclides are stable, apparently living forever. Unstable nuclides decay (that is, they are radioactive), eventually producing a stable nuclide after many decays. We call the original nuclide the **parent** and its decay products the **daughters**. Some radioactive nuclides decay in a single step to a stable nucleus. For example, ^{60}Co is unstable and decays directly to ^{60}Ni , which is stable. Others, such as ^{238}U , decay to another unstable nuclide, resulting in a **decay series** in which each subsequent nuclide decays until a stable nuclide is finally produced. The decay series that starts from ^{238}U is of particular interest, since it produces the radioactive isotopes ^{226}Ra and ^{210}Po , which the Curies first discovered (see [\[link\]](#)). Radon gas is also produced (^{222}Rn in the series), an increasingly recognized naturally occurring hazard. Since radon is a noble gas, it emanates from materials, such as soil, containing even trace amounts of ^{238}U and can be inhaled. The decay of radon and its daughters produces internal damage. The ^{238}U decay series ends with ^{206}Pb , a stable isotope of lead.



The decay series produced by ^{238}U , the most common uranium isotope.

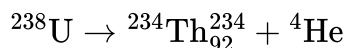
Nuclides are graphed in the same manner as in the chart of nuclides. The type of decay for each member of the series is shown, as well as the half-lives. Note that some nuclides decay by more than one mode. You can see why radium and polonium are found in uranium ore. A stable isotope of lead is the end product of the series.

Note that the daughters of α decay shown in [\[link\]](#) always have two fewer protons and two fewer neutrons than the parent. This seems reasonable, since we know that α decay is the emission of a ^4He nucleus, which has two protons and two neutrons. The daughters of β decay have one less neutron and one more proton than their parent. Beta decay is a little more subtle, as we shall see. No γ decays are shown in the figure, because they do not produce a daughter that differs from the parent.

Alpha Decay

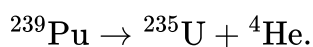
In **alpha decay**, a ${}^4\text{He}$ nucleus simply breaks away from the parent nucleus, leaving a daughter with two fewer protons and two fewer neutrons than the parent (see [\[link\]](#)). One example of α decay is shown in [\[link\]](#) for ${}^{238}\text{U}$. Another nuclide that undergoes α decay is ${}^{239}\text{Pu}$. The decay equations for these two nuclides are

Equation:



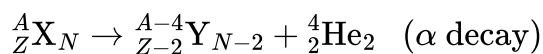
and

Equation:



Alpha decay is the separation of a ${}^4\text{He}$ nucleus from the parent. The daughter nucleus has two fewer protons and two fewer neutrons than the parent. Alpha decay occurs spontaneously only if the daughter and ${}^4\text{He}$ nucleus have less total mass than the parent.

If you examine the periodic table of the elements, you will find that Th has $Z = 90$, two fewer than U, which has $Z = 92$. Similarly, in the second **decay equation**, we see that U has two fewer protons than Pu, which has $Z = 94$. The general rule for α decay is best written in the format ${}^A_Z\text{X}_N$. If a certain nuclide is known to α decay (generally this information must be looked up in a table of isotopes, such as in [Appendix B](#)), its α **decay equation** is

Equation:

where Y is the nuclide that has two fewer protons than X, such as Th having two fewer than U. So if you were told that ${}^{239}\text{Pu}$ α decays and were asked to write the complete decay equation, you would first look up which element has two fewer protons (an atomic number two lower) and find that this is uranium. Then since four nucleons have broken away from the original 239, its atomic mass would be 235.

It is instructive to examine conservation laws related to α decay. You can see from the equation ${}_Z^AX_N \rightarrow {}_{Z-2}^{A-4}Y_{N-2} + {}_2^4\text{He}_2$ that total charge is conserved. Linear and angular momentum are conserved, too. Although conserved angular momentum is not of great consequence in this type of decay, conservation of linear momentum has interesting consequences. If the nucleus is at rest when it decays, its momentum is zero. In that case, the fragments must fly in opposite directions with equal-magnitude momenta so that total momentum remains zero. This results in the α particle carrying away most of the energy, as a bullet from a heavy rifle carries away most of the energy of the powder burned to shoot it. Total mass–energy is also conserved: the energy produced in the decay comes from conversion of a fraction of the original mass. As discussed in [Atomic Physics](#), the general relationship is

Equation:

$$E = (\Delta m)c^2.$$

Here, E is the **nuclear reaction energy** (the reaction can be nuclear decay or any other reaction), and Δm is the difference in mass between initial and final products. When the final products have less total mass, Δm is positive, and the reaction releases energy (is exothermic). When the products have greater total mass, the reaction is endothermic (Δm is negative) and must be induced with an energy input. For α decay to be spontaneous, the decay products must have smaller mass than the parent.

Example:**Alpha Decay Energy Found from Nuclear Masses**

Find the energy emitted in the α decay of ${}^{239}\text{Pu}$.

Strategy

Nuclear reaction energy, such as released in α decay, can be found using the equation $E = (\Delta m)c^2$. We must first find Δm , the difference in mass between the parent nucleus and the products of the decay. This is easily done using masses given in [Appendix A](#).

Solution

The decay equation was given earlier for ${}^{239}\text{Pu}$; it is

Equation:



Thus the pertinent masses are those of ^{239}Pu , ^{235}U , and the α particle or ^4He , all of which are listed in [Appendix A](#). The initial mass was $m(^{239}\text{Pu}) = 239.052157 \text{ u}$. The final mass is the sum $m(^{235}\text{U}) + m(^4\text{He}) = 235.043924 \text{ u} + 4.002602 \text{ u} = 239.046526 \text{ u}$. Thus,

Equation:

$$\begin{aligned}\Delta m &= m(^{239}\text{Pu}) - [m(^{235}\text{U}) + m(^4\text{He})] \\ &= 239.052157 \text{ u} - 239.046526 \text{ u} \\ &= 0.005631 \text{ u}.\end{aligned}$$

Now we can find E by entering Δm into the equation:

Equation:

$$E = (\Delta m)c^2 = (0.005631 \text{ u})c^2.$$

We know $1 \text{ u} = 931.5 \text{ MeV}/c^2$, and so

Equation:

$$E = (0.005631)(931.5 \text{ MeV}/c^2)(c^2) = 5.25 \text{ MeV}.$$

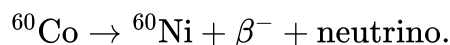
Discussion

The energy released in this α decay is in the MeV range, about 10^6 times as great as typical chemical reaction energies, consistent with many previous discussions. Most of this energy becomes kinetic energy of the α particle (or ^4He nucleus), which moves away at high speed. The energy carried away by the recoil of the ^{235}U nucleus is much smaller in order to conserve momentum. The ^{235}U nucleus can be left in an excited state to later emit photons (γ rays). This decay is spontaneous and releases energy, because the products have less mass than the parent nucleus. The question of why the products have less mass will be discussed in [Binding Energy](#). Note that the masses given in [Appendix A](#) are atomic masses of neutral atoms, including their electrons. The mass of the electrons is the same before and after α decay, and so their masses subtract out when finding Δm . In this case, there are 94 electrons before and after the decay.

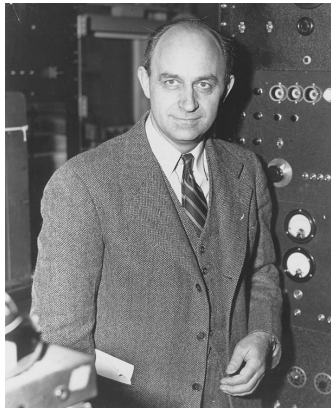
Beta Decay

There are actually *three* types of **beta decay**. The first discovered was “ordinary” beta decay and is called β^- decay or electron emission. The symbol β^- represents *an electron emitted in nuclear beta decay*. Cobalt-60 is a nuclide that β^- decays in the following manner:

Equation:



The **neutrino** is a particle emitted in beta decay that was unanticipated and is of fundamental importance. The neutrino was not even proposed in theory until more than 20 years after beta decay was known to involve electron emissions. Neutrinos are so difficult to detect that the first direct evidence of them was not obtained until 1953. Neutrinos are nearly massless, have no charge, and do not interact with nucleons via the strong nuclear force. Traveling approximately at the speed of light, they have little time to affect any nucleus they encounter. This is, owing to the fact that they have no charge (and they are not EM waves), they do not interact through the EM force. They do interact via the relatively weak and very short range weak nuclear force. Consequently, neutrinos escape almost any detector and penetrate almost any shielding. However, neutrinos do carry energy, angular momentum (they are fermions with half-integral spin), and linear momentum away from a beta decay. When accurate measurements of beta decay were made, it became apparent that energy, angular momentum, and linear momentum were not accounted for by the daughter nucleus and electron alone. Either a previously unsuspected particle was carrying them away, or three conservation laws were being violated. Wolfgang Pauli made a formal proposal for the existence of neutrinos in 1930. The Italian-born American physicist Enrico Fermi (1901–1954) gave neutrinos their name, meaning little neutral ones, when he developed a sophisticated theory of beta decay (see [link](#)). Part of Fermi's theory was the identification of the weak nuclear force as being distinct from the strong nuclear force and in fact responsible for beta decay.



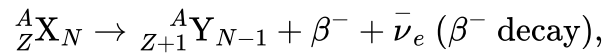
Enrico Fermi was
nearly unique
among 20th-
century physicists
—he made
significant
contributions both
as an
experimentalist and
a theorist. His
many contributions
to theoretical

physics included the identification of the weak nuclear force. The fermi (fm) is named after him, as are an entire class of subatomic particles (fermions), an element (Fermium), and a major research laboratory (Fermilab). His experimental work included studies of radioactivity, for which he won the 1938 Nobel Prize in physics, and creation of the first nuclear chain reaction. (credit: United States Department of Energy, Office of Public Affairs)

The neutrino also reveals a new conservation law. There are various families of particles, one of which is the electron family. We propose that the number of members of the electron family is constant in any process or any closed system. In our example of beta decay, there are no members of the electron family present before the decay, but after, there is an electron and a neutrino. So electrons are given an electron family number of +1. The neutrino in β^- decay is an **electron's antineutrino**, given the symbol $\bar{\nu}_e$, where ν is the Greek letter nu, and the subscript e means this neutrino is related to the electron. The bar indicates this is a particle of **antimatter**. (All particles have antimatter counterparts that are nearly identical except that they have the opposite charge. Antimatter is almost entirely absent on Earth, but it is found in nuclear decay and other nuclear and particle reactions as well as in outer space.) The electron's antineutrino $\bar{\nu}_e$, being antimatter, has an electron family number of -1. The total is zero, before and after the decay. The new conservation law, obeyed in all circumstances, states that the *total electron family number is constant*. An electron cannot be created without also creating an antimatter family member. This law is analogous to the conservation of charge in a situation where total charge is originally zero, and equal amounts of positive and negative charge must be created in a reaction to keep the total zero.

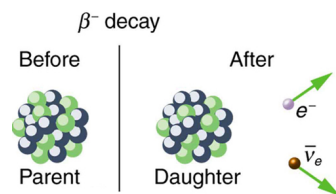
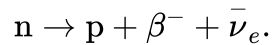
If a nuclide ${}^A_Z\text{X}_N$ is known to β^- decay, then its β^- decay equation is

Equation:



where Y is the nuclide having one more proton than X (see [\[link\]](#)). So if you know that a certain nuclide β^- decays, you can find the daughter nucleus by first looking up Z for the parent and then determining which element has atomic number $Z + 1$. In the example of the β^- decay of ${}^{60}\text{Co}$ given earlier, we see that $Z = 27$ for Co and $Z = 28$ is Ni. It is as if one of the neutrons in the parent nucleus decays into a proton, electron, and neutrino. In fact, neutrons outside of nuclei do just that—they live only an average of a few minutes and β^- decay in the following manner:

Equation:



In β^- decay, the parent nucleus emits an electron and an antineutrino.

The daughter nucleus has one more proton and one less neutron than its parent.

Neutrinos interact so weakly that they are almost never directly observed, but they play a fundamental role in particle physics.

We see that charge is conserved in β^- decay, since the total charge is Z before and after the decay. For example, in ^{60}Co decay, total charge is 27 before decay, since cobalt has $Z = 27$. After decay, the daughter nucleus is Ni, which has $Z = 28$, and there is an electron, so that the total charge is also $28 + (-1)$ or 27. Angular momentum is conserved, but not obviously (you have to examine the spins and angular momenta of the final products in detail to verify this). Linear momentum is also conserved, again imparting most of the decay energy to the electron and the antineutrino, since they are of low and zero mass, respectively. Another new conservation law is obeyed here and elsewhere in nature. *The total number of nucleons A is conserved.* In ^{60}Co decay, for example, there are 60 nucleons before and after the decay. Note that total A is also conserved in α decay. Also note that the total number of protons changes, as does the total number of neutrons, so that total Z and total N are *not* conserved in β^- decay, as they are in α decay. Energy released in β^- decay can be calculated given the masses of the parent and products.

Example:

β^- Decay Energy from Masses

Find the energy emitted in the β^- decay of ^{60}Co .

Strategy and Concept

As in the preceding example, we must first find Δm , the difference in mass between the parent nucleus and the products of the decay, using masses given in [Appendix A](#). Then the emitted energy is calculated as before, using $E = (\Delta m)c^2$. The initial mass is just that of the parent nucleus, and the final mass is that of the daughter nucleus and the electron created in the decay. The neutrino is massless, or nearly so. However, since the masses given in [Appendix A](#) are for neutral atoms, the daughter nucleus has one more electron than the parent, and so the extra electron mass that corresponds to the β^- is included in the atomic mass of Ni. Thus,

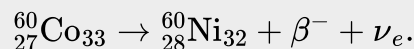
Equation:

$$\Delta m = m(^{60}\text{Co}) - m(^{60}\text{Ni}).$$

Solution

The β^- decay equation for ^{60}Co is

Equation:



As noticed,

Equation:

$$\Delta m = m(^{60}\text{Co}) - m(^{60}\text{Ni}).$$

Entering the masses found in [Appendix A](#) gives

Equation:

$$\Delta m = 59.933820 \text{ u} - 59.930789 \text{ u} = 0.003031 \text{ u}.$$

Thus,

Equation:

$$E = (\Delta m)c^2 = (0.003031 \text{ u})c^2.$$

Using $1 \text{ u} = 931.5 \text{ MeV}/c^2$, we obtain

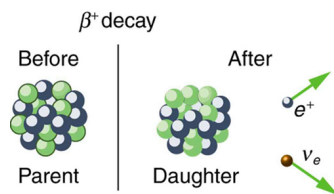
Equation:

$$E = (0.003031)(931.5 \text{ MeV}/c^2)(c^2) = 2.82 \text{ MeV}.$$

Discussion and Implications

Perhaps the most difficult thing about this example is convincing yourself that the β^- mass is included in the atomic mass of ^{60}Ni . Beyond that are other implications. Again the decay energy is in the MeV range. This energy is shared by all of the products of the decay. In many ^{60}Co decays, the daughter nucleus ^{60}Ni is left in an excited state and emits photons (γ rays). Most of the remaining energy goes to the electron and neutrino, since the recoil kinetic energy of the daughter nucleus is small. One final note: the electron emitted in β^- decay is created in the nucleus at the time of decay.

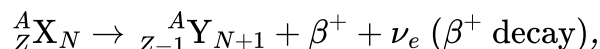
The second type of beta decay is less common than the first. It is β^+ decay. Certain nuclides decay by the emission of a *positive* electron. This is **antielectron** or **positron decay** (see [\[link\]](#)).



β^+ decay is the emission of a positron that eventually finds an electron to annihilate, characteristically producing gammas in opposite directions.

The antielectron is often represented by the symbol e^+ , but in beta decay it is written as β^+ to indicate the antielectron was emitted in a nuclear decay. Antielectrons are the antimatter counterpart to electrons, being nearly identical, having the same mass, spin, and so on, but having a positive charge and an electron family number of -1 . When a **positron** encounters an electron, there is a mutual annihilation in which all the mass of the antielectron-electron pair is converted into pure photon energy. (The reaction, $e^+ + e^- \rightarrow \gamma + \gamma$, conserves electron family number as well as all other conserved quantities.) If a nuclide ${}^A_Z\text{X}_N$ is known to β^+ decay, then its β^+ **decay equation** is

Equation:



where Y is the nuclide having one less proton than X (to conserve charge) and ν_e is the symbol for the **electron's neutrino**, which has an electron family number of $+1$. Since an antimatter member of the electron family (the β^+) is created in the decay, a matter member of the family (here the ν_e) must also be created. Given, for example, that ${}^{22}\text{Na}$ β^+ decays, you can write its full decay equation by first finding that $Z = 11$ for ${}^{22}\text{Na}$, so that the daughter nuclide will have $Z = 10$, the atomic number for neon. Thus the β^+ decay equation for ${}^{22}\text{Na}$ is

Equation:



In β^+ decay, it is as if one of the protons in the parent nucleus decays into a neutron, a positron, and a neutrino. Protons do not do this outside of the nucleus, and so the decay is due to the complexities of the nuclear force. Note again that the total number of nucleons is constant in this and any other reaction. To find the energy emitted in β^+ decay, you must again count the number of electrons in the neutral atoms, since atomic masses are used. The daughter has one less electron than the parent, and one electron mass is created in the decay. Thus, in β^+ decay,

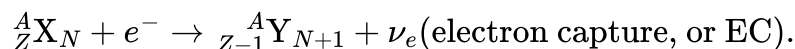
Equation:

$$\Delta m = m(\text{parent}) - [m(\text{daughter}) + 2m_e],$$

since we use the masses of neutral atoms.

Electron capture is the third type of beta decay. Here, a nucleus captures an inner-shell electron and undergoes a nuclear reaction that has the same effect as β^+ decay. Electron capture is sometimes denoted by the letters EC. We know that electrons cannot reside in the nucleus, but this is a nuclear reaction that consumes the electron and occurs spontaneously only when the products have less mass than the parent plus the electron. If a nuclide ${}^A_Z\text{X}_N$ is known to undergo electron capture, then its **electron capture equation** is

Equation:



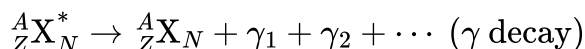
Any nuclide that can β^+ decay can also undergo electron capture (and often does both). The same conservation laws are obeyed for EC as for β^+ decay. It is good practice to confirm these for yourself.

All forms of beta decay occur because the parent nuclide is unstable and lies outside the region of stability in the chart of nuclides. Those nuclides that have relatively more neutrons than those in the region of stability will β^- decay to produce a daughter with fewer neutrons, producing a daughter nearer the region of stability. Similarly, those nuclides having relatively more protons than those in the region of stability will β^+ decay or undergo electron capture to produce a daughter with fewer protons, nearer the region of stability.

Gamma Decay

Gamma decay is the simplest form of nuclear decay—it is the emission of energetic photons by nuclei left in an excited state by some earlier process. Protons and neutrons in an excited nucleus are in higher orbitals, and they fall to lower levels by photon emission (analogous to electrons in excited atoms). Nuclear excited states have lifetimes typically of only about 10^{-14} s, an indication of the great strength of the forces pulling the nucleons to lower states. The γ decay equation is simply

Equation:



where the asterisk indicates the nucleus is in an excited state. There may be one or more γ s emitted, depending on how the nuclide de-excites. In radioactive decay, γ emission is common and is preceded by γ or β decay. For example, when ${}^{60}\text{Co}$ β^- decays, it most often leaves the daughter nucleus in an excited state, written ${}^{60}\text{Ni}^*$. Then the nickel nucleus quickly γ decays by the emission of two penetrating γ s:

Equation:



These are called cobalt γ rays, although they come from nickel—they are used for cancer therapy, for example. It is again constructive to verify the conservation laws for gamma decay. Finally, since γ decay does not change the nuclide to another species, it is not prominently featured in charts of decay series, such as that in [\[link\]](#).

There are other types of nuclear decay, but they occur less commonly than α , β , and γ decay. Spontaneous fission is the most important of the other forms of nuclear decay because of its applications in nuclear power and weapons. It is covered in the next chapter.

Section Summary

- When a parent nucleus decays, it produces a daughter nucleus following rules and conservation laws. There are three major types of nuclear decay, called alpha (α), beta (β), and gamma (γ). The α decay equation is

Equation:

$${}^A_Z\text{X}_N \rightarrow {}^{A-4}_{Z-2}\text{Y}_{N-2} + {}^4_2\text{He}_2.$$

- Nuclear decay releases an amount of energy E related to the mass destroyed Δm by

Equation:

$$E = (\Delta m)c^2.$$

- There are three forms of beta decay. The β^- decay equation is

Equation:

$${}^A_Z\text{X}_N \rightarrow {}^A_{Z+1}\text{Y}_{N-1} + \beta^- + \nu_e.$$

- The β^+ decay equation is

Equation:

$${}^A_Z\text{X}_N \rightarrow {}^A_{Z-1}\text{Y}_{N+1} + \beta^+ + \nu_e.$$

- The electron capture equation is

Equation:

$${}^A_Z\text{X}_N + e^- \rightarrow {}^A_{Z-1}\text{Y}_{N+1} + \nu_e.$$

- β^- is an electron, β^+ is an antielectron or positron, ν_e represents an electron's neutrino, and $\bar{\nu}_e$ is an electron's antineutrino. In addition to all previously known conservation laws, two new ones arise— conservation of electron family number and conservation of the total number of nucleons. The γ decay equation is

Equation:

$${}^A_Z\text{X}_N^* \rightarrow {}^A_Z\text{X}_N + \gamma_1 + \gamma_2 + \cdots$$

γ is a high-energy photon originating in a nucleus.

Conceptual Questions

Exercise:

Problem:

Star Trek fans have often heard the term “antimatter drive.” Describe how you could use a magnetic field to trap antimatter, such as produced by nuclear decay, and later combine it with matter to produce energy. Be specific about the type of antimatter, the need for vacuum storage, and the fraction of matter converted into energy.

Exercise:**Problem:**

What conservation law requires an electron’s neutrino to be produced in electron capture? Note that the electron no longer exists after it is captured by the nucleus.

Exercise:**Problem:**

Neutrinos are experimentally determined to have an extremely small mass. Huge numbers of neutrinos are created in a supernova at the same time as massive amounts of light are first produced. When the 1987A supernova occurred in the Large Magellanic Cloud, visible primarily in the Southern Hemisphere and some 100,000 light-years away from Earth, neutrinos from the explosion were observed at about the same time as the light from the blast. How could the relative arrival times of neutrinos and light be used to place limits on the mass of neutrinos?

Exercise:**Problem:**

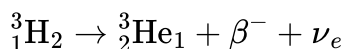
What do the three types of beta decay have in common that is distinctly different from alpha decay?

Problems & Exercises

In the following eight problems, write the complete decay equation for the given nuclide in the complete ${}^A_Z\text{X}_N$ notation. Refer to the periodic table for values of Z .

Exercise:**Problem:**

β^- decay of ${}^3\text{H}$ (tritium), a manufactured isotope of hydrogen used in some digital watch displays, and manufactured primarily for use in hydrogen bombs.

Solution:**Equation:**

Exercise:

Problem:

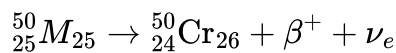
β^- decay of ^{40}K , a naturally occurring rare isotope of potassium responsible for some of our exposure to background radiation.

Exercise:

Problem: β^+ decay of ^{50}Mn .

Solution:

Equation:



Exercise:

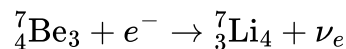
Problem: β^+ decay of ^{52}Fe .

Exercise:

Problem: Electron capture by ^7Be .

Solution:

Equation:



Exercise:

Problem: Electron capture by ^{106}In .

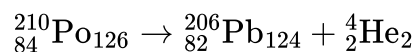
Exercise:

Problem:

α decay of ^{210}Po , the isotope of polonium in the decay series of ^{238}U that was discovered by the Curies. A favorite isotope in physics labs, since it has a short half-life and decays to a stable nuclide.

Solution:

Equation:



Exercise:**Problem:**

α decay of ^{226}Ra , another isotope in the decay series of ^{238}U , first recognized as a new element by the Curies. Poses special problems because its daughter is a radioactive noble gas.

In the following four problems, identify the parent nuclide and write the complete decay equation in the ${}^A_Z\text{X}_N$ notation. Refer to the periodic table for values of Z .

Exercise:**Problem:**

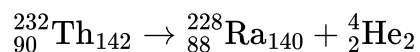
β^- decay producing ^{137}Ba . The parent nuclide is a major waste product of reactors and has chemistry similar to potassium and sodium, resulting in its concentration in your cells if ingested.

Solution:**Equation:****Exercise:****Problem:**

β^- decay producing ^{90}Y . The parent nuclide is a major waste product of reactors and has chemistry similar to calcium, so that it is concentrated in bones if ingested (^{90}Y is also radioactive.)

Exercise:**Problem:**

α decay producing ^{228}Ra . The parent nuclide is nearly 100% of the natural element and is found in gas lantern mantles and in metal alloys used in jets (^{228}Ra is also radioactive).

Solution:**Equation:****Exercise:**

Problem:

α decay producing ^{208}Pb . The parent nuclide is in the decay series produced by ^{232}Th , the only naturally occurring isotope of thorium.

Exercise:**Problem:**

When an electron and positron annihilate, both their masses are destroyed, creating two equal energy photons to preserve momentum. (a) Confirm that the annihilation equation $e^+ + e^- \rightarrow \gamma + \gamma$ conserves charge, electron family number, and total number of nucleons. To do this, identify the values of each before and after the annihilation. (b) Find the energy of each γ ray, assuming the electron and positron are initially nearly at rest. (c) Explain why the two γ rays travel in exactly opposite directions if the center of mass of the electron-positron system is initially at rest.

Solution:

(a)

charge: $(+1) + (-1) = 0$; electron family number: $(+1) + (-1) = 0$; A : $0 + 0 = 0$

(b) 0.511 MeV

(c) The two γ rays must travel in exactly opposite directions in order to conserve momentum, since initially there is zero momentum if the center of mass is initially at rest.

Exercise:**Problem:**

Confirm that charge, electron family number, and the total number of nucleons are all conserved by the rule for α decay given in the equation ${}_Z^AX_N \rightarrow {}_{Z-2}^{A-4}Y_{N-2} + {}_2^4\text{He}_2$. To do this, identify the values of each before and after the decay.

Exercise:**Problem:**

Confirm that charge, electron family number, and the total number of nucleons are all conserved by the rule for β^- decay given in the equation ${}_Z^AX_N \rightarrow {}_{Z+1}^AY_{N-1} + \beta^- + \nu_e$. To do this, identify the values of each before and after the decay.

Solution:**Equation:**

$$Z = (Z + 1) - 1; \quad A = A; \quad \text{efn} : 0 = (+1) + (-1)$$

Exercise:**Problem:**

Confirm that charge, electron family number, and the total number of nucleons are all conserved by the rule for β^- decay given in the equation ${}^A_Z\text{X}_N \rightarrow {}^A_{Z-1}\text{Y}_{N-1} + \beta^- + \nu_e$. To do this, identify the values of each before and after the decay.

Exercise:**Problem:**

Confirm that charge, electron family number, and the total number of nucleons are all conserved by the rule for electron capture given in the equation ${}^A_Z\text{X}_N + e^- \rightarrow {}^A_{Z-1}\text{Y}_{N+1} + \nu_e$. To do this, identify the values of each before and after the capture.

Solution:**Equation:**

$$Z - 1 = Z - 1; \quad A = A; \quad \text{efn} : (+1) = (+1)$$

Exercise:**Problem:**

A rare decay mode has been observed in which ${}^{222}\text{Ra}$ emits a ${}^{14}\text{C}$ nucleus. (a) The decay equation is ${}^{222}\text{Ra} \rightarrow {}^A\text{X} + {}^{14}\text{C}$. Identify the nuclide ${}^A\text{X}$. (b) Find the energy emitted in the decay. The mass of ${}^{222}\text{Ra}$ is 222.015353 u.

Exercise:

Problem: (a) Write the complete α decay equation for ${}^{226}\text{Ra}$.

(b) Find the energy released in the decay.

Solution:

(a) ${}^{226}_{88}\text{Ra}_{138} \rightarrow {}^{222}_{86}\text{Rn}_{136} + {}^4_2\text{He}_2$

(b) 4.87 MeV

Exercise:

Problem: (a) Write the complete α decay equation for ${}^{249}\text{Cf}$.

(b) Find the energy released in the decay.

Exercise:**Problem:**

(a) Write the complete β^- decay equation for the neutron. (b) Find the energy released in the decay.

Solution:

(a) $n \rightarrow p + \beta^- + \nu_e$

(b)) 0.783 MeV

Exercise:**Problem:**

(a) Write the complete β^- decay equation for ^{90}Sr , a major waste product of nuclear reactors. (b) Find the energy released in the decay.

Exercise:**Problem:**

Calculate the energy released in the β^+ decay of ^{22}Na , the equation for which is given in the text. The masses of ^{22}Na and ^{22}Ne are 21.994434 and 21.991383 u, respectively.

Solution:

1.82 MeV

Exercise:

Problem: (a) Write the complete β^+ decay equation for ^{11}C .

(b) Calculate the energy released in the decay. The masses of ^{11}C and ^{11}B are 11.011433 and 11.009305 u, respectively.

Exercise:

Problem: (a) Calculate the energy released in the α decay of ^{238}U .

(b) What fraction of the mass of a single ^{238}U is destroyed in the decay? The mass of ^{234}Th is 234.043593 u.

(c) Although the fractional mass loss is large for a single nucleus, it is difficult to observe for an entire macroscopic sample of uranium. Why is this?

Solution:

(a) 4.274 MeV

(b) 1.927×10^{-5}

(c) Since U-238 is a slowly decaying substance, only a very small number of nuclei decay on human timescales; therefore, although those nuclei that decay lose a noticeable fraction of their mass, the change in the total mass of the sample is not detectable for a macroscopic sample.

Exercise:

Problem: (a) Write the complete reaction equation for electron capture by ${}^7\text{Be}$.

(b) Calculate the energy released.

Exercise:

Problem: (a) Write the complete reaction equation for electron capture by ${}^{15}\text{O}$.

(b) Calculate the energy released.

Solution:

(a) ${}^{15}_8\text{O}_7 + e^- \rightarrow {}^{15}_7\text{N}_8 + \nu_e$

(b) 2.754 MeV

Glossary

parent

the original state of nucleus before decay

daughter

the nucleus obtained when parent nucleus decays and produces another nucleus following the rules and the conservation laws

positron

the particle that results from positive beta decay; also known as an antielectron

decay

the process by which an atomic nucleus of an unstable atom loses mass and energy by emitting ionizing particles

alpha decay

type of radioactive decay in which an atomic nucleus emits an alpha particle

beta decay

type of radioactive decay in which an atomic nucleus emits a beta particle

gamma decay

type of radioactive decay in which an atomic nucleus emits a gamma particle

decay equation

the equation to find out how much of a radioactive material is left after a given period of time

nuclear reaction energy

the energy created in a nuclear reaction

neutrino

an electrically neutral, weakly interacting elementary subatomic particle

electron's antineutrino

antiparticle of electron's neutrino

positron decay

type of beta decay in which a proton is converted to a neutron, releasing a positron and a neutrino

antielectron

another term for positron

decay series

process whereby subsequent nuclides decay until a stable nuclide is produced

electron's neutrino

a subatomic elementary particle which has no net electric charge

antimatter

composed of antiparticles

electron capture

the process in which a proton-rich nuclide absorbs an inner atomic electron and simultaneously emits a neutrino

electron capture equation

equation representing the electron capture

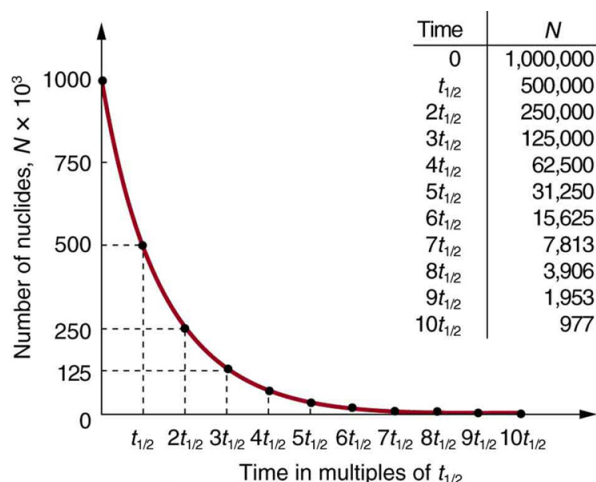
Half-Life and Activity

- Define half-life.
- Define dating.
- Calculate age of old objects by radioactive dating.

Unstable nuclei decay. However, some nuclides decay faster than others. For example, radium and polonium, discovered by the Curies, decay faster than uranium. This means they have shorter lifetimes, producing a greater rate of decay. In this section we explore half-life and activity, the quantitative terms for lifetime and rate of decay.

Half-Life

Why use a term like half-life rather than lifetime? The answer can be found by examining [\[link\]](#), which shows how the number of radioactive nuclei in a sample decreases with time. The *time in which half of the original number of nuclei decay* is defined as the **half-life**, $t_{1/2}$. Half of the remaining nuclei decay in the next half-life. Further, half of that amount decays in the following half-life. Therefore, the number of radioactive nuclei decreases from N to $N/2$ in one half-life, then to $N/4$ in the next, and to $N/8$ in the next, and so on. If N is a large number, then *many* half-lives (not just two) pass before all of the nuclei decay. Nuclear decay is an example of a purely statistical process. A more precise definition of half-life is that *each nucleus has a 50% chance of living for a time equal to one half-life $t_{1/2}$* . Thus, if N is reasonably large, half of the original nuclei decay in a time of one half-life. If an individual nucleus makes it through that time, it still has a 50% chance of surviving through another half-life. Even if it happens to make it through hundreds of half-lives, it still has a 50% chance of surviving through one more. The probability of decay is the same no matter when you start counting. This is like random coin flipping. The chance of heads is 50%, no matter what has happened before.



Radioactive decay reduces the number of radioactive nuclei over time. In one half-life $t_{1/2}$, the number decreases to half of its original value. Half of what remains decay in the next half-life, and half of those in the next, and so on. This is an exponential decay, as seen in the graph of the number of nuclei present as a function of time.

There is a tremendous range in the half-lives of various nuclides, from as short as 10^{-23} s for the most unstable, to more than 10^{16} y for the least unstable, or about 46 orders of magnitude. Nuclides with the shortest half-lives are those for which the nuclear forces are least attractive, an indication of the extent to which the nuclear force can depend on the particular combination of neutrons and protons. The concept of half-life is applicable to other subatomic particles, as will be discussed in [Particle Physics](#). It is also applicable to the decay of excited states in atoms and nuclei. The following equation gives the quantitative relationship between the original

number of nuclei present at time zero (N_0) and the number (N) at a later time t :

Equation:

$$N = N_0 e^{-\lambda t},$$

where $e = 2.71828\dots$ is the base of the natural logarithm, and λ is the **decay constant** for the nuclide. The shorter the half-life, the larger is the value of λ , and the faster the exponential $e^{-\lambda t}$ decreases with time. The relationship between the decay constant λ and the half-life $t_{1/2}$ is

Equation:

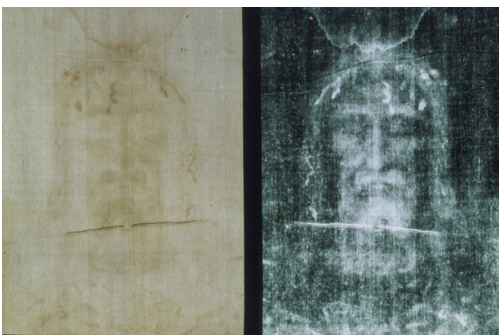
$$\lambda = \frac{\ln(2)}{t_{1/2}} \approx \frac{0.693}{t_{1/2}}.$$

To see how the number of nuclei declines to half its original value in one half-life, let $t = t_{1/2}$ in the exponential in the equation $N = N_0 e^{-\lambda t}$. This gives $N = N_0 e^{-\lambda t} = N_0 e^{-0.693} = 0.500 N_0$. For integral numbers of half-lives, you can just divide the original number by 2 over and over again, rather than using the exponential relationship. For example, if ten half-lives have passed, we divide N by 2 ten times. This reduces it to $N/1024$. For an arbitrary time, not just a multiple of the half-life, the exponential relationship must be used.

Radioactive dating is a clever use of naturally occurring radioactivity. Its most famous application is **carbon-14 dating**. Carbon-14 has a half-life of 5730 years and is produced in a nuclear reaction induced when solar neutrinos strike ^{14}N in the atmosphere. Radioactive carbon has the same chemistry as stable carbon, and so it mixes into the ecosphere, where it is consumed and becomes part of every living organism. Carbon-14 has an abundance of 1.3 parts per trillion of normal carbon. Thus, if you know the number of carbon nuclei in an object (perhaps determined by mass and Avogadro's number), you multiply that number by 1.3×10^{-12} to find the number of ^{14}C nuclei in the object. When an organism dies, carbon exchange with the environment ceases, and ^{14}C is not replenished as it

decays. By comparing the abundance of ^{14}C in an artifact, such as mummy wrappings, with the normal abundance in living tissue, it is possible to determine the artifact's age (or time since death). Carbon-14 dating can be used for biological tissues as old as 50 or 60 thousand years, but is most accurate for younger samples, since the abundance of ^{14}C nuclei in them is greater. Very old biological materials contain no ^{14}C at all. There are instances in which the date of an artifact can be determined by other means, such as historical knowledge or tree-ring counting. These cross-references have confirmed the validity of carbon-14 dating and permitted us to calibrate the technique as well. Carbon-14 dating revolutionized parts of archaeology and is of such importance that it earned the 1960 Nobel Prize in chemistry for its developer, the American chemist Willard Libby (1908–1980).

One of the most famous cases of carbon-14 dating involves the Shroud of Turin, a long piece of fabric purported to be the burial shroud of Jesus (see [\[link\]](#)). This relic was first displayed in Turin in 1354 and was denounced as a fraud at that time by a French bishop. Its remarkable negative imprint of an apparently crucified body resembles the then-accepted image of Jesus, and so the shroud was never disregarded completely and remained controversial over the centuries. Carbon-14 dating was not performed on the shroud until 1988, when the process had been refined to the point where only a small amount of material needed to be destroyed. Samples were tested at three independent laboratories, each being given four pieces of cloth, with only one unidentified piece from the shroud, to avoid prejudice. All three laboratories found samples of the shroud contain 92% of the ^{14}C found in living tissues, allowing the shroud to be dated (see [\[link\]](#)).



Part of the Shroud of Turin, which shows a remarkable negative imprint likeness of Jesus complete with evidence of crucifixion wounds. The shroud first surfaced in the 14th century and was only recently carbon-14 dated. It has not been determined how the image was placed on the material. (credit: Butko, Wikimedia Commons)

Example:**How Old Is the Shroud of Turin?**

Calculate the age of the Shroud of Turin given that the amount of ^{14}C found in it is 92% of that in living tissue.

Strategy

Knowing that 92% of the ^{14}C remains means that $N/N_0 = 0.92$.

Therefore, the equation $N = N_0 e^{-\lambda t}$ can be used to find λt . We also know that the half-life of ^{14}C is 5730 y, and so once λt is known, we can use the equation $\lambda = \frac{0.693}{t_{1/2}}$ to find λ and then find t as requested. Here, we postulate that the decrease in ^{14}C is solely due to nuclear decay.

Solution

Solving the equation $N = N_0 e^{-\lambda t}$ for N/N_0 gives

Equation:

$$\frac{N}{N_0} = e^{-\lambda t}.$$

Thus,

Equation:

$$0.92 = e^{-\lambda t}.$$

Taking the natural logarithm of both sides of the equation yields

Equation:

$$\ln 0.92 = -\lambda t$$

so that

Equation:

$$-0.0834 = -\lambda t.$$

Rearranging to isolate t gives

Equation:

$$t = \frac{0.0834}{\lambda}.$$

Now, the equation $\lambda = \frac{0.693}{t_{1/2}}$ can be used to find λ for ^{14}C . Solving for λ and substituting the known half-life gives

Equation:

$$\lambda = \frac{0.693}{t_{1/2}} = \frac{0.693}{5730 \text{ y}}.$$

We enter this value into the previous equation to find t :

Equation:

$$t = \frac{0.0834}{\frac{0.693}{5730 \text{ y}}} = 690 \text{ y}.$$

Discussion

This dates the material in the shroud to $1988 - 690 = \text{a.d. } 1300$. Our calculation is only accurate to two digits, so that the year is rounded to 1300. The values obtained at the three independent laboratories gave a

weighted average date of a.d. 1320 ± 60 . The uncertainty is typical of carbon-14 dating and is due to the small amount of ^{14}C in living tissues, the amount of material available, and experimental uncertainties (reduced by having three independent measurements). It is meaningful that the date of the shroud is consistent with the first record of its existence and inconsistent with the period in which Jesus lived.

There are other forms of radioactive dating. Rocks, for example, can sometimes be dated based on the decay of ^{238}U . The decay series for ^{238}U ends with ^{206}Pb , so that the ratio of these nuclides in a rock is an indication of how long it has been since the rock solidified. The original composition of the rock, such as the absence of lead, must be known with some confidence. However, as with carbon-14 dating, the technique can be verified by a consistent body of knowledge. Since ^{238}U has a half-life of 4.5×10^9 y, it is useful for dating only very old materials, showing, for example, that the oldest rocks on Earth solidified about 3.5×10^9 years ago.

Activity, the Rate of Decay

What do we mean when we say a source is highly radioactive? Generally, this means the number of decays per unit time is very high. We define **activity** R to be the **rate of decay** expressed in decays per unit time. In equation form, this is

Equation:

$$R = \frac{\Delta N}{\Delta t}$$

where ΔN is the number of decays that occur in time Δt . The SI unit for activity is one decay per second and is given the name **becquerel** (Bq) in honor of the discoverer of radioactivity. That is,

Equation:

$$1 \text{ Bq} = 1 \text{ decay/s.}$$

Activity R is often expressed in other units, such as decays per minute or decays per year. One of the most common units for activity is the **curie** (Ci), defined to be the activity of 1 g of ^{226}Ra , in honor of Marie Curie's work with radium. The definition of curie is

Equation:

$$1 \text{ Ci} = 3.70 \times 10^{10} \text{ Bq,}$$

or 3.70×10^{10} decays per second. A curie is a large unit of activity, while a becquerel is a relatively small unit. $1 \text{ MBq} = 100$ microcuries (μCi). In countries like Australia and New Zealand that adhere more to SI units, most radioactive sources, such as those used in medical diagnostics or in physics laboratories, are labeled in Bq or megabecquerel (MBq).

Intuitively, you would expect the activity of a source to depend on two things: the amount of the radioactive substance present, and its half-life. The greater the number of radioactive nuclei present in the sample, the more will decay per unit of time. The shorter the half-life, the more decays per unit time, for a given number of nuclei. So activity R should be proportional to the number of radioactive nuclei, N , and inversely proportional to their half-life, $t_{1/2}$. In fact, your intuition is correct. It can be shown that the activity of a source is

Equation:

$$R = \frac{0.693N}{t_{1/2}}$$

where N is the number of radioactive nuclei present, having half-life $t_{1/2}$. This relationship is useful in a variety of calculations, as the next two examples illustrate.

Example:**How Great Is the ^{14}C Activity in Living Tissue?**

Calculate the activity due to ^{14}C in 1.00 kg of carbon found in a living organism. Express the activity in units of Bq and Ci.

Strategy

To find the activity R using the equation $R = \frac{0.693N}{t_{1/2}}$, we must know N and $t_{1/2}$. The half-life of ^{14}C can be found in [Appendix B](#), and was stated above as 5730 y. To find N , we first find the number of ^{12}C nuclei in 1.00 kg of carbon using the concept of a mole. As indicated, we then multiply by 1.3×10^{-12} (the abundance of ^{14}C in a carbon sample from a living organism) to get the number of ^{14}C nuclei in a living organism.

Solution

One mole of carbon has a mass of 12.0 g, since it is nearly pure ^{12}C . (A mole has a mass in grams equal in magnitude to A found in the periodic table.) Thus the number of carbon nuclei in a kilogram is

Equation:

$$N(^{12}\text{C}) = \frac{6.02 \times 10^{23} \text{ mol}^{-1}}{12.0 \text{ g/mol}} \times (1000 \text{ g}) = 5.02 \times 10^{25}.$$

So the number of ^{14}C nuclei in 1 kg of carbon is

Equation:

$$N(^{14}\text{C}) = (5.02 \times 10^{25})(1.3 \times 10^{-12}) = 6.52 \times 10^{13}.$$

Now the activity R is found using the equation $R = \frac{0.693N}{t_{1/2}}$.

Entering known values gives

Equation:

$$R = \frac{0.693(6.52 \times 10^{13})}{5730 \text{ y}} = 7.89 \times 10^9 \text{ y}^{-1},$$

or 7.89×10^9 decays per year. To convert this to the unit Bq, we simply convert years to seconds. Thus,

Equation:

$$R = (7.89 \times 10^9 \text{ y}^{-1}) \frac{1.00 \text{ y}}{3.16 \times 10^7 \text{ s}} = 250 \text{ Bq},$$

or 250 decays per second. To express R in curies, we use the definition of a curie,

Equation:

$$R = \frac{250 \text{ Bq}}{3.7 \times 10^{10} \text{ Bq/Ci}} = 6.76 \times 10^{-9} \text{ Ci}.$$

Thus,

Equation:

$$R = 6.76 \text{ nCi}.$$

Discussion

Our own bodies contain kilograms of carbon, and it is intriguing to think there are hundreds of ^{14}C decays per second taking place in us. Carbon-14 and other naturally occurring radioactive substances in our bodies contribute to the background radiation we receive. The small number of decays per second found for a kilogram of carbon in this example gives you some idea of how difficult it is to detect ^{14}C in a small sample of material. If there are 250 decays per second in a kilogram, then there are 0.25 decays per second in a gram of carbon in living tissue. To observe this, you must be able to distinguish decays from other forms of radiation, in order to reduce background noise. This becomes more difficult with an old tissue sample, since it contains less ^{14}C , and for samples more than 50 thousand years old, it is impossible.

Human-made (or artificial) radioactivity has been produced for decades and has many uses. Some of these include medical therapy for cancer, medical imaging and diagnostics, and food preservation by irradiation. Many applications as well as the biological effects of radiation are explored in [Medical Applications of Nuclear Physics](#), but it is clear that radiation is hazardous. A number of tragic examples of this exist, one of the most disastrous being the meltdown and fire at the Chernobyl reactor complex in

the Ukraine (see [\[link\]](#)). Several radioactive isotopes were released in huge quantities, contaminating many thousands of square kilometers and directly affecting hundreds of thousands of people. The most significant releases were of ^{131}I , ^{90}Sr , ^{137}Cs , ^{239}Pu , ^{238}U , and ^{235}U . Estimates are that the total amount of radiation released was about 100 million curies.

Human and Medical Applications



The Chernobyl reactor.
More than 100 people
died soon after its
meltdown, and there will
be thousands of deaths
from radiation-induced
cancer in the future.

While the accident was
due to a series of human
errors, the cleanup efforts
were heroic. Most of the
immediate fatalities were
firefighters and reactor
personnel. (credit: Elena
Filatova)

Example:**What Mass of ^{137}Cs Escaped Chernobyl?**

It is estimated that the Chernobyl disaster released 6.0 MCi of ^{137}Cs into the environment. Calculate the mass of ^{137}Cs released.

Strategy

We can calculate the mass released using Avogadro's number and the concept of a mole if we can first find the number of nuclei N released.

Since the activity R is given, and the half-life of ^{137}Cs is found in [Appendix B](#) to be 30.2 y, we can use the equation $R = \frac{0.693N}{t_{1/2}}$ to find N .

Solution

Solving the equation $R = \frac{0.693N}{t_{1/2}}$ for N gives

Equation:

$$N = \frac{Rt_{1/2}}{0.693}.$$

Entering the given values yields

Equation:

$$N = \frac{(6.0 \text{ MCi})(30.2 \text{ y})}{0.693}.$$

Converting curies to becquerels and years to seconds, we get

Equation:

$$\begin{aligned} N &= \frac{(6.0 \times 10^6 \text{ Ci})(3.7 \times 10^{10} \text{ Bq/Ci})(30.2 \text{ y})(3.16 \times 10^7 \text{ s/y})}{0.693} \\ &= 3.1 \times 10^{26}. \end{aligned}$$

One mole of a nuclide $^A X$ has a mass of A grams, so that one mole of ^{137}Cs has a mass of 137 g. A mole has 6.02×10^{23} nuclei. Thus the mass of ^{137}Cs released was

Equation:

$$\begin{aligned} m &= \left(\frac{137 \text{ g}}{6.02 \times 10^{23}} \right) (3.1 \times 10^{26}) = 70 \times 10^3 \text{ g} \\ &= 70 \text{ kg}. \end{aligned}$$

Discussion

While 70 kg of material may not be a very large mass compared to the amount of fuel in a power plant, it is extremely radioactive, since it only has a 30-year half-life. Six megacuries (6.0 MCi) is an extraordinary amount of activity but is only a fraction of what is produced in nuclear reactors. Similar amounts of the other isotopes were also released at Chernobyl. Although the chances of such a disaster may have seemed small, the consequences were extremely severe, requiring greater caution than was used. More will be said about safe reactor design in the next chapter, but it should be noted that Western reactors have a fundamentally safer design.

Activity R decreases in time, going to half its original value in one half-life, then to one-fourth its original value in the next half-life, and so on. Since $R = \frac{0.693N}{t_{1/2}}$, the activity decreases as the number of radioactive nuclei decreases. The equation for R as a function of time is found by combining the equations $N = N_0 e^{-\lambda t}$ and $R = \frac{0.693N}{t_{1/2}}$, yielding

Equation:

$$R = R_0 e^{-\lambda t},$$

where R_0 is the activity at $t = 0$. This equation shows exponential decay of radioactive nuclei. For example, if a source originally has a 1.00-mCi activity, it declines to 0.500 mCi in one half-life, to 0.250 mCi in two half-lives, to 0.125 mCi in three half-lives, and so on. For times other than whole half-lives, the equation $R = R_0 e^{-\lambda t}$ must be used to find R .

Note:

PhET Explorations: Alpha Decay

Watch alpha particles escape from a polonium nucleus, causing radioactive alpha decay. See how random decay times relate to the half life.

Section Summary

- Half-life $t_{1/2}$ is the time in which there is a 50% chance that a nucleus will decay. The number of nuclei N as a function of time is

Equation:

$$N = N_0 e^{-\lambda t},$$

where N_0 is the number present at $t = 0$, and λ is the decay constant, related to the half-life by

Equation:

$$\lambda = \frac{0.693}{t_{1/2}}.$$

- One of the applications of radioactive decay is radioactive dating, in which the age of a material is determined by the amount of radioactive decay that occurs. The rate of decay is called the activity R :

Equation:

$$R = \frac{\Delta N}{\Delta t}.$$

- The SI unit for R is the becquerel (Bq), defined by
- Equation:**

$$1 \text{ Bq} = 1 \text{ decay/s}.$$

- R is also expressed in terms of curies (Ci), where

Equation:

$$1 \text{ Ci} = 3.70 \times 10^{10} \text{ Bq}.$$

- The activity R of a source is related to N and $t_{1/2}$ by

Equation:

$$R = \frac{0.693N}{t_{1/2}}.$$

- Since N has an exponential behavior as in the equation $N = N_0 e^{-\lambda t}$, the activity also has an exponential behavior, given by

Equation:

$$R = R_0 e^{-\lambda t},$$

where R_0 is the activity at $t = 0$.

Conceptual Questions

Exercise:

Problem:

In a 3×10^9 -year-old rock that originally contained some ^{238}U , which has a half-life of 4.5×10^9 years, we expect to find some ^{238}U remaining in it. Why are ^{226}Ra , ^{222}Rn , and ^{210}Po also found in such a rock, even though they have much shorter half-lives (1600 years, 3.8 days, and 138 days, respectively)?

Exercise:

Problem:

Does the number of radioactive nuclei in a sample decrease to *exactly* half its original value in one half-life? Explain in terms of the statistical nature of radioactive decay.

Exercise:

Problem:

Radioactivity depends on the nucleus and not the atom or its chemical state. Why, then, is one kilogram of uranium more radioactive than one kilogram of uranium hexafluoride?

Exercise:**Problem:**

Explain how a bound system can have less mass than its components. Why is this not observed classically, say for a building made of bricks?

Exercise:**Problem:**

Spontaneous radioactive decay occurs only when the decay products have less mass than the parent, and it tends to produce a daughter that is more stable than the parent. Explain how this is related to the fact that more tightly bound nuclei are more stable. (Consider the binding energy per nucleon.)

Exercise:**Problem:**

To obtain the most precise value of BE from the equation $BE = [ZM(^1\text{H}) + Nm_n]c^2 - m(^A\text{X})c^2$, we should take into account the binding energy of the electrons in the neutral atoms. Will doing this produce a larger or smaller value for BE? Why is this effect usually negligible?

Exercise:**Problem:**

How does the finite range of the nuclear force relate to the fact that BE/A is greatest for A near 60?

Problems & Exercises

Data from the appendices and the periodic table may be needed for these problems.

Exercise:

Problem:

An old campfire is uncovered during an archaeological dig. Its charcoal is found to contain less than 1/1000 the normal amount of ^{14}C . Estimate the minimum age of the charcoal, noting that $2^{10} = 1024$.

Solution:

57,300 y

Exercise:

Problem:

A ^{60}Co source is labeled 4.00 mCi, but its present activity is found to be 1.85×10^7 Bq. (a) What is the present activity in mCi? (b) How long ago did it actually have a 4.00-mCi activity?

Exercise:

Problem:

(a) Calculate the activity R in curies of 1.00 g of ^{226}Ra . (b) Discuss why your answer is not exactly 1.00 Ci, given that the curie was originally supposed to be exactly the activity of a gram of radium.

Solution:

(a) 0.988 Ci

(b) The half-life of ^{226}Ra is now better known.

Exercise:

Problem:

Show that the activity of the ^{14}C in 1.00 g of ^{12}C found in living tissue is 0.250 Bq.

Exercise:**Problem:**

Mantles for gas lanterns contain thorium, because it forms an oxide that can survive being heated to incandescence for long periods of time. Natural thorium is almost 100% ^{232}Th , with a half-life of 1.405×10^{10} y. If an average lantern mantle contains 300 mg of thorium, what is its activity?

Solution:

$$1.22 \times 10^3 \text{ Bq}$$

Exercise:**Problem:**

Cow's milk produced near nuclear reactors can be tested for as little as 1.00 pCi of ^{131}I per liter, to check for possible reactor leakage. What mass of ^{131}I has this activity?

Exercise:**Problem:**

(a) Natural potassium contains ^{40}K , which has a half-life of 1.277×10^9 y. What mass of ^{40}K in a person would have a decay rate of 4140 Bq? (b) What is the fraction of ^{40}K in natural potassium, given that the person has 140 g in his body? (These numbers are typical for a 70-kg adult.)

Solution:

(a) 16.0 mg

(b) 0.0114%

Exercise:

Problem:

There is more than one isotope of natural uranium. If a researcher isolates 1.00 mg of the relatively scarce ^{235}U and finds this mass to have an activity of 80.0 Bq, what is its half-life in years?

Exercise:

Problem:

^{50}V has one of the longest known radioactive half-lives. In a difficult experiment, a researcher found that the activity of 1.00 kg of ^{50}V is 1.75 Bq. What is the half-life in years?

Solution:

$$1.48 \times 10^{17} \text{ y}$$

Exercise:

Problem:

You can sometimes find deep red crystal vases in antique stores, called uranium glass because their color was produced by doping the glass with uranium. Look up the natural isotopes of uranium and their half-lives, and calculate the activity of such a vase assuming it has 2.00 g of uranium in it. Neglect the activity of any daughter nuclides.

Exercise:

Problem:

A tree falls in a forest. How many years must pass before the ^{14}C activity in 1.00 g of the tree's carbon drops to 1.00 decay per hour?

Solution:

$$5.6 \times 10^4 \text{ y}$$

Exercise:**Problem:**

What fraction of the ^{40}K that was on Earth when it formed 4.5×10^9 years ago is left today?

Exercise:**Problem:**

A 5000-Ci ^{60}Co source used for cancer therapy is considered too weak to be useful when its activity falls to 3500 Ci. How long after its manufacture does this happen?

Solution:

2.71 y

Exercise:**Problem:**

Natural uranium is 0.7200% ^{235}U and 99.27% ^{238}U . What were the percentages of ^{235}U and ^{238}U in natural uranium when Earth formed 4.5×10^9 years ago?

Exercise:**Problem:**

The β^- particles emitted in the decay of ^3H (tritium) interact with matter to create light in a glow-in-the-dark exit sign. At the time of manufacture, such a sign contains 15.0 Ci of ^3H . (a) What is the mass of the tritium? (b) What is its activity 5.00 y after manufacture?

Solution:

(a) 1.56 mg

(b) 11.3 Ci

Exercise:**Problem:**

World War II aircraft had instruments with glowing radium-painted dials (see [\[link\]](#)). The activity of one such instrument was 1.0×10^5 Bq when new. (a) What mass of ^{226}Ra was present? (b) After some years, the phosphors on the dials deteriorated chemically, but the radium did not escape. What is the activity of this instrument 57.0 years after it was made?

Exercise:**Problem:**

(a) The ^{210}Po source used in a physics laboratory is labeled as having an activity of $1.0 \mu\text{Ci}$ on the date it was prepared. A student measures the radioactivity of this source with a Geiger counter and observes 1500 counts per minute. She notices that the source was prepared 120 days before her lab. What fraction of the decays is she observing with her apparatus? (b) Identify some of the reasons that only a fraction of the α s emitted are observed by the detector.

Solution:

(a) 1.23×10^{-3}

(b) Only part of the emitted radiation goes in the direction of the detector. Only a fraction of that causes a response in the detector. Some of the emitted radiation (mostly α particles) is observed within the source. Some is absorbed within the source, some is absorbed by the detector, and some does not penetrate the detector.

Exercise:

Problem:

Armor-piercing shells with depleted uranium cores are fired by aircraft at tanks. (The high density of the uranium makes them effective.) The uranium is called depleted because it has had its ^{235}U removed for reactor use and is nearly pure ^{238}U . Depleted uranium has been erroneously called non-radioactive. To demonstrate that this is wrong: (a) Calculate the activity of 60.0 g of pure ^{238}U . (b) Calculate the activity of 60.0 g of natural uranium, neglecting the ^{234}U and all daughter nuclides.

Exercise:**Problem:**

The ceramic glaze on a red-orange Fiestaware plate is U_2O_3 and contains 50.0 grams of ^{238}U , but very little ^{235}U . (a) What is the activity of the plate? (b) Calculate the total energy that will be released by the ^{238}U decay. (c) If energy is worth 12.0 cents per $\text{kW} \cdot \text{h}$, what is the monetary value of the energy emitted? (These plates went out of production some 30 years ago, but are still available as collectibles.)

Solution:

(a) $1.68 \times 10^{-5} \text{ Ci}$

(b) $8.65 \times 10^{10} \text{ J}$

(c) $\$2.9 \times 10^3$

Exercise:

Problem:

Large amounts of depleted uranium (^{238}U) are available as a by-product of uranium processing for reactor fuel and weapons. Uranium is very dense and makes good counter weights for aircraft. Suppose you have a 4000-kg block of ^{238}U . (a) Find its activity. (b) How many calories per day are generated by thermalization of the decay energy? (c) Do you think you could detect this as heat? Explain.

Exercise:**Problem:**

The *Galileo* space probe was launched on its long journey past several planets in 1989, with an ultimate goal of Jupiter. Its power source is 11.0 kg of ^{238}Pu , a by-product of nuclear weapons plutonium production. Electrical energy is generated thermoelectrically from the heat produced when the 5.59-MeV α particles emitted in each decay crash to a halt inside the plutonium and its shielding. The half-life of ^{238}Pu is 87.7 years. (a) What was the original activity of the ^{238}Pu in becquerel? (b) What power was emitted in kilowatts? (c) What power was emitted 12.0 y after launch? You may neglect any extra energy from daughter nuclides and any losses from escaping γ rays.

Solution:

(a) $6.97 \times 10^{15} \text{ Bq}$

(b) 6.24 kW

(c) 5.67 kW

Exercise:**Problem: Construct Your Own Problem**

Consider the generation of electricity by a radioactive isotope in a space probe, such as described in [\[link\]](#). Construct a problem in which you calculate the mass of a radioactive isotope you need in order to

supply power for a long space flight. Among the things to consider are the isotope chosen, its half-life and decay energy, the power needs of the probe and the length of the flight.

Exercise:

Problem: Unreasonable Results

A nuclear physicist finds $1.0\ \mu\text{g}$ of ^{236}U in a piece of uranium ore and assumes it is primordial since its half-life is $2.3 \times 10^7\ \text{y}$. (a) Calculate the amount of ^{236}U that would have had to have been on Earth when it formed $4.5 \times 10^9\ \text{y}$ ago for $1.0\ \mu\text{g}$ to be left today. (b) What is unreasonable about this result? (c) What assumption is responsible?

Exercise:

Problem: Unreasonable Results

(a) Repeat [\[link\]](#) but include the 0.0055% natural abundance of ^{234}U with its $2.45 \times 10^5\ \text{y}$ half-life. (b) What is unreasonable about this result? (c) What assumption is responsible? (d) Where does the ^{234}U come from if it is not primordial?

Exercise:

Problem: Unreasonable Results

The manufacturer of a smoke alarm decides that the smallest current of α radiation he can detect is $1.00\ \mu\text{A}$. (a) Find the activity in curies of an α emitter that produces a $1.00\ \mu\text{A}$ current of α particles. (b) What is unreasonable about this result? (c) What assumption is responsible?

Solution:

(a) $84.5\ \text{Ci}$

(b) An extremely large activity, many orders of magnitude greater than permitted for home use.

(c) The assumption of $1.00\ \mu\text{A}$ is unreasonably large. Other methods can detect much smaller decay rates.

Glossary

becquerel

SI unit for rate of decay of a radioactive material

half-life

the time in which there is a 50% chance that a nucleus will decay

radioactive dating

an application of radioactive decay in which the age of a material is determined by the amount of radioactivity of a particular type that occurs

decay constant

quantity that is inversely proportional to the half-life and that is used in equation for number of nuclei as a function of time

carbon-14 dating

a radioactive dating technique based on the radioactivity of carbon-14

activity

the rate of decay for radioactive nuclides

rate of decay

the number of radioactive events per unit time

curie

the activity of 1g of ^{226}Ra , equal to $3.70 \times 10^{10}\ \text{Bq}$

Atomic Masses

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
0	neutron	1	n	1.008 665	β^-	10.37 min
1	Hydrogen	1	^1H	1.007 825	99.985%	
	Deuterium	2	^2H or D	2.014 102	0.015%	
	Tritium	3	^3H or T	3.016 050	β^-	12.33 y
2	Helium	3	^3He	3.016 030	$1.38 \times 10^{-4}\%$	
		4	^4He	4.002 603	$\approx 100\%$	
3	Lithium	6	^6Li	6.015 121	7.5%	
		7	^7Li	7.016 003	92.5%	
4	Beryllium	7	^7Be	7.016 928	EC	53.29 d
		9	^9Be	9.012 182	100%	
5	Boron	10	^{10}B	10.012 937	19.9%	
		11	^{11}B	11.009 305	80.1%	
6	Carbon	11	^{11}C	11.011 432	EC, β^+	
		12	^{12}C	12.000 000	98.90%	
		13	^{13}C	13.003 355	1.10%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
		14	^{14}C	14.003 241	β^-	5730 y
7	Nitrogen	13	^{13}N	13.005 738	β^+	9.96 min
		14	^{14}N	14.003 074	99.63%	
		15	^{15}N	15.000 108	0.37%	
8	Oxygen	15	^{15}O	15.003 065	EC, β^+	122 s
		16	^{16}O	15.994 915	99.76%	
		18	^{18}O	17.999 160	0.200%	
9	Fluorine	18	^{18}F	18.000 937	EC, β^+	1.83 h
		19	^{19}F	18.998 403	100%	
10	Neon	20	^{20}Ne	19.992 435	90.51%	
		22	^{22}Ne	21.991 383	9.22%	
11	Sodium	22	^{22}Na	21.994 434	β^+	2.602 y
		23	^{23}Na	22.989 767	100%	
		24	^{24}Na	23.990 961	β^-	14.96 h
12	Magnesium	24	^{24}Mg	23.985 042	78.99%	
13	Aluminum	27	^{27}Al	26.981 539	100%	
14	Silicon	28	^{28}Si	27.976 927	92.23%	2.62h

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
		31	^{31}Si	30.975 362	β^-	
15	Phosphorus	31	^{31}P	30.973 762	100%	
		32	^{32}P	31.973 907	β^-	14.28 d
16	Sulfur	32	^{32}S	31.972 070	95.02%	
		35	^{35}S	34.969 031	β^-	87.4 d
17	Chlorine	35	^{35}Cl	34.968 852	75.77%	
		37	^{37}Cl	36.965 903	24.23%	
18	Argon	40	^{40}Ar	39.962 384	99.60%	
19	Potassium	39	^{39}K	38.963 707	93.26%	
		40	^{40}K	39.963 999	0.0117%, EC, β^-	$1.28 \times 10^9 \text{ y}$
20	Calcium	40	^{40}Ca	39.962 591	96.94%	
21	Scandium	45	^{45}Sc	44.955 910	100%	
22	Titanium	48	^{48}Ti	47.947 947	73.8%	
23	Vanadium	51	^{51}V	50.943 962	99.75%	
24	Chromium	52	^{52}Cr	51.940 509	83.79%	
25	Manganese	55	^{55}Mn	54.938 047	100%	
26	Iron	56	^{56}Fe	55.934 939	91.72%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
27	Cobalt	59	^{59}Co	58.933 198	100%	
		60	^{60}Co	59.933 819	β^-	5.271 y
28	Nickel	58	^{58}Ni	57.935 346	68.27%	
		60	^{60}Ni	59.930 788	26.10%	
29	Copper	63	^{63}Cu	62.939 598	69.17%	
		65	^{65}Cu	64.927 793	30.83%	
30	Zinc	64	^{64}Zn	63.929 145	48.6%	
		66	^{66}Zn	65.926 034	27.9%	
31	Gallium	69	^{69}Ga	68.925 580	60.1%	
32	Germanium	72	^{72}Ge	71.922 079	27.4%	
		74	^{74}Ge	73.921 177	36.5%	
33	Arsenic	75	^{75}As	74.921 594	100%	
34	Selenium	80	^{80}Se	79.916 520	49.7%	
35	Bromine	79	^{79}Br	78.918 336	50.69%	
36	Krypton	84	^{84}Kr	83.911 507	57.0%	
37	Rubidium	85	^{85}Rb	84.911 794	72.17%	
38	Strontium	86	^{86}Sr	85.909 267	9.86%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
		88	^{88}Sr	87.905 619	82.58%	
		90	^{90}Sr	89.907 738	β^-	28.8 y
39	Yttrium	89	^{89}Y	88.905 849	100%	
		90	^{90}Y	89.907 152	β^-	64.1 h
40	Zirconium	90	^{90}Zr	89.904 703	51.45%	
41	Niobium	93	^{93}Nb	92.906 377	100%	
42	Molybdenum	98	^{98}Mo	97.905 406	24.13%	
43	Technetium	98	^{98}Tc	97.907 215	β^-	$4.2 \times 10^6 \text{ y}$
44	Ruthenium	102	^{102}Ru	101.904 348	31.6%	
45	Rhodium	103	^{103}Rh	102.905 500	100%	
46	Palladium	106	^{106}Pd	105.903 478	27.33%	
47	Silver	107	^{107}Ag	106.905 092	51.84%	
		109	^{109}Ag	108.904 757	48.16%	
48	Cadmium	114	^{114}Cd	113.903 357	28.73%	
49	Indium	115	^{115}In	114.903 880	95.7%, β^-	$4.4 \times 10^{14} \text{ y}$
50	Tin	120	^{120}Sn	119.902 200	32.59%	
51	Antimony	121	^{121}Sb	120.903 821	57.3%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
52	Tellurium	130	^{130}Te	129.906 229	33.8%, β^-	$2.5 \times 10^{21}\text{y}$
53	Iodine	127	^{127}I	126.904 473	100%	
		131	^{131}I	130.906 114	β^-	8.040 d
54	Xenon	132	^{132}Xe	131.904 144	26.9%	
		136	^{136}Xe	135.907 214	8.9%	
55	Cesium	133	^{133}Cs	132.905 429	100%	
		134	^{134}Cs	133.906 696	EC, β^-	2.06 y
56	Barium	137	^{137}Ba	136.905 812	11.23%	
		138	^{138}Ba	137.905 232	71.70%	
57	Lanthanum	139	^{139}La	138.906 346	99.91%	
58	Cerium	140	^{140}Ce	139.905 433	88.48%	
59	Praseodymium	141	^{141}Pr	140.907 647	100%	
60	Neodymium	142	^{142}Nd	141.907 719	27.13%	
61	Promethium	145	^{145}Pm	144.912 743	EC, α	17.7 y
62	Samarium	152	^{152}Sm	151.919 729	26.7%	
63	Europium	153	^{153}Eu	152.921 225	52.2%	
64	Gadolinium	158	^{158}Gd	157.924 099	24.84%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
65	Terbium	159	^{159}Tb	158.925 342	100%	
66	Dysprosium	164	^{164}Dy	163.929 171	28.2%	
67	Holmium	165	^{165}Ho	164.930 319	100%	
68	Erbium	166	^{166}Er	165.930 290	33.6%	
69	Thulium	169	^{169}Tm	168.934 212	100%	
70	Ytterbium	174	^{174}Yb	173.938 859	31.8%	
71	Lutecium	175	^{175}Lu	174.940 770	97.41%	
72	Hafnium	180	^{180}Hf	179.946 545	35.10%	
73	Tantalum	181	^{181}Ta	180.947 992	99.98%	
74	Tungsten	184	^{184}W	183.950 928	30.67%	
75	Rhenium	187	^{187}Re	186.955 744	62.6%, β^-	$4.6 \times 10^{10}\text{y}$
76	Osmium	191	^{191}Os	190.960 920	β^-	15.4 d
		192	^{192}Os	191.961 467	41.0%	
77	Iridium	191	^{191}Ir	190.960 584	37.3%	
		193	^{193}Ir	192.962 917	62.7%	
78	Platinum	195	^{195}Pt	194.964 766	33.8%	
79	Gold	197	^{197}Au	196.966 543	100%	

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
		198	^{198}Au	197.968 217	β^-	2.696 d
80	Mercury	199	^{199}Hg	198.968 253	16.87%	
		202	^{202}Hg	201.970 617	29.86%	
81	Thallium	205	^{205}Tl	204.974 401	70.48%	
82	Lead	206	^{206}Pb	205.974 440	24.1%	
		207	^{207}Pb	206.975 872	22.1%	
		208	^{208}Pb	207.976 627	52.4%	
		210	^{210}Pb	209.984 163	α, β^-	22.3 y
		211	^{211}Pb	210.988 735	β^-	36.1 min
		212	^{212}Pb	211.991 871	β^-	10.64 h
83	Bismuth	209	^{209}Bi	208.980 374	100%	
		211	^{211}Bi	210.987 255	α, β^-	2.14 min
84	Polonium	210	^{210}Po	209.982 848	α	138.38 d
85	Astatine	218	^{218}At	218.008 684	α, β^-	1.6 s
86	Radon	222	^{222}Rn	222.017 570	α	3.82 d
87	Francium	223	^{223}Fr	223.019 733	α, β^-	21.8 min
88	Radium	226	^{226}Ra	226.025 402	α	$1.60 \times 10^3 \text{ y}$

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
89	Actinium	227	^{227}Ac	227.027 750	α, β^-	21.8 y
90	Thorium	228	^{228}Th	228.028 715	α	1.91 y
		232	^{232}Th	232.038 054	100%, α	$1.41 \times 10^{10}\text{y}$
91	Protactinium	231	^{231}Pa	231.035 880	α	$3.28 \times 10^4\text{y}$
92	Uranium	233	^{233}U	233.039 628	α	$1.59 \times 10^3\text{y}$
		235	^{235}U	235.043 924	0.720%, α	$7.04 \times 10^8\text{y}$
		236	^{236}U	236.045 562	α	$2.34 \times 10^7\text{y}$
		238	^{238}U	238.050 784	99.2745%, α	$4.47 \times 10^9\text{y}$
		239	^{239}U	239.054 289	β^-	23.5 min
93	Neptunium	239	^{239}Np	239.052 933	β^-	2.355 d
94	Plutonium	239	^{239}Pu	239.052 157	α	$2.41 \times 10^4\text{y}$
95	Americium	243	^{243}Am	243.061 375	α , fission	$7.37 \times 10^3\text{y}$
96	Curium	245	^{245}Cm	245.065 483	α	$8.50 \times 10^3\text{y}$
97	Berkelium	247	^{247}Bk	247.070 300	α	$1.38 \times 10^3\text{y}$
98	Californium	249	^{249}Cf	249.074 844	α	351 y
99	Einsteinium	254	^{254}Es	254.088 019	α, β^-	276 d
100	Fermium	253	^{253}Fm	253.085 173	EC, α	3.00 d

Atomic Number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
101	Mendelevium	255	^{255}Md	255.091 081	EC, α	27 min
102	Nobelium	255	^{255}No	255.093 260	EC, α	3.1 min
103	Lawrencium	257	^{257}Lr	257.099 480	EC, α	0.646 s
104	Rutherfordium	261	^{261}Rf	261.108 690	α	1.08 min
105	Dubnium	262	^{262}Db	262.113 760	α , fission	34 s
106	Seaborgium	263	^{263}Sg	263.11 86	α , fission	0.8 s
107	Bohrium	262	^{262}Bh	262.123 1	α	0.102 s
108	Hassium	264	^{264}Hs	264.128 5	α	0.08 ms
109	Meitnerium	266	^{266}Mt	266.137 8	α	3.4 ms

Atomic Masses

Selected Radioactive Isotopes

Decay modes are α , β^- , β^+ , electron capture (EC) and isomeric transition (IT). EC results in the same daughter nucleus as would β^+ decay. IT is a transition from a metastable excited state. Energies for β^\pm decays are the maxima; average energies are roughly one-half the maxima.

Isotope	$t_{1/2}$	DecayMode(s)	Energy(MeV)	Percent		γ -Ray Energy(MeV)
^3H	12.33 y	β^-	0.0186	100%		
^{14}C	5730 y	β^-	0.156	100%		
^{13}N	9.96 min	β^+	1.20	100%		
^{22}Na	2.602 y	β^+	0.55	90%	γ	1.27
^{32}P	14.28 d	β^-	1.71	100%		
^{35}S	87.4 d	β^-	0.167	100%		
^{36}Cl	$3.00 \times 10^5 \text{ y}$	β^-	0.710	100%		
^{40}K	$1.28 \times 10^9 \text{ y}$	β^-	1.31	89%		
^{43}K	22.3 h	β^-	0.827	87%	$\gamma \text{ s}$	0.373
						0.618
^{45}Ca	165 d	β^-	0.257	100%		
^{51}Cr	27.70 d	EC			γ	0.320
^{52}Mn	5.59d	β^+	3.69	28%	$\gamma \text{ s}$	1.33
						1.43
^{52}Fe	8.27 h	β^+	1.80	43%		0.169
						0.378
^{59}Fe	44.6 d	$\beta^- \text{ s}$	0.273	45%	$\gamma \text{ s}$	1.10
			0.466	55%		1.29
^{60}Co	5.271 y	β^-	0.318	100%	$\gamma \text{ s}$	1.17
						1.33
^{65}Zn	244.1 d	EC			γ	1.12

Isotope	$t_{1/2}$	DecayMode(s)	Energy(MeV)	Percent		γ -Ray Energy(MeV)
^{67}Ga	78.3 h	EC			γ s	0.0933
						0.185
						0.300
						others
^{75}Se	118.5 d	EC			γ s	0.121
						0.136
						0.265
						0.280
						others
^{86}Rb	18.8 d	β^- s	0.69	9%	γ	1.08
			1.77	91%		
^{85}Sr	64.8 d	EC			γ	0.514
^{90}Sr	28.8 y	β^-	0.546	100%		
^{90}Y	64.1 h	β^-	2.28	100%		
$^{99\text{m}}\text{Tc}$	6.02 h	IT			γ	0.142
$^{113\text{m}}\text{In}$	99.5 min	IT			γ	0.392
^{123}I	13.0 h	EC			γ	0.159
^{131}I	8.040 d	β^- s	0.248	7%	γ s	0.364
			0.607	93%		others
			others			
^{129}Cs	32.3 h	EC			γ s	0.0400
						0.372
						0.411
						others
^{137}Cs	30.17 y	β^- s	0.511	95%	γ	0.662
			1.17	5%		

Isotope	$t_{1/2}$	DecayMode(s)	Energy(MeV)	Percent		γ -Ray Energy(MeV)
^{140}Ba	12.79 d	β^-	1.035	$\approx 100\%$	γ s	0.030
						0.044
						0.537
						others
^{198}Au	2.696 d	β^-	1.161	$\approx 100\%$	γ	0.412
^{197}Hg	64.1 h	EC			γ	0.0733
^{210}Po	138.38 d	α	5.41	100%		
^{226}Ra	$1.60 \times 10^3 \text{ y}$	α s	4.68	5%	γ	0.186
			4.87	95%		
^{235}U	$7.038 \times 10^8 \text{ y}$	α	4.68	$\approx 100\%$	γ s	numerous
^{238}U	$4.468 \times 10^9 \text{ y}$	α s	4.22	23%	γ	0.050
			4.27	77%		
^{237}Np	$2.14 \times 10^6 \text{ y}$	α s	numerous		γ s	numerous
			4.96 (max.)			
^{239}Pu	$2.41 \times 10^4 \text{ y}$	α s	5.19	11%	γ s	7.5×10^{-5}
			5.23	15%		0.013
			5.24	73%		0.052
						others
^{243}Am	$7.37 \times 10^3 \text{ y}$	α s	Max. 5.44		γ s	0.075
			5.37	88%		others
			5.32	11%		
			others			

Selected Radioactive Isotopes

Useful Information

This appendix is broken into several tables.

- [\[link\]](#), Important Constants
- [\[link\]](#), Submicroscopic Masses
- [\[link\]](#), Solar System Data
- [\[link\]](#), Metric Prefixes for Powers of Ten and Their Symbols
- [\[link\]](#), The Greek Alphabet
- [\[link\]](#), SI units
- [\[link\]](#), Selected British Units
- [\[link\]](#), Other Units
- [\[link\]](#), Useful Formulae

Symbol	Meaning	Best Value	Approximate Value
c	Speed of light in vacuum	$2.99792458 \times 10^8 \text{ m/s}$	$3.00 \times 10^8 \text{ m/s}$
G	Gravitational constant	$6.67408(31) \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$	$6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$
N_A	Avogadro's number	$6.02214129(27) \times 10^{23}$	6.02×10^{23}
k	Boltzmann's constant	$1.3806488(13) \times 10^{-23} \text{ J/K}$	$1.38 \times 10^{-23} \text{ J/K}$
R	Gas constant	$8.3144621(75) \text{ J/mol} \cdot \text{K}$	$8.31 \text{ J/mol} \cdot \text{K} = 1.99 \text{ cal/mol} \cdot \text{K} =$
σ	Stefan-Boltzmann constant	$5.670373(21) \times 10^{-8} \text{ W/m}^2 \cdot \text{K}$	$5.67 \times 10^{-8} \text{ W/m}^2 \cdot \text{K}$
k	Coulomb force constant	$8.987551788... \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$	$8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$
q_e	Charge on electron	$-1.602176565(35) \times 10^{-19} \text{ C}$	$-1.60 \times 10^{-19} \text{ C}$
ϵ_0	Permittivity of free space	$8.854187817... \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2$	$8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2$
μ_0	Permeability of free space	$4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$	$1.26 \times 10^{-6} \text{ T} \cdot \text{m/A}$
h	Planck's constant	$6.62606957(29) \times 10^{-34} \text{ J} \cdot \text{s}$	$6.63 \times 10^{-34} \text{ J} \cdot \text{s}$

Important Constants^[footnote]

Stated values are according to the National Institute of Standards and Technology Reference on Constants, Units, and Uncertainty, www.physics.nist.gov/cuu (accessed May 18, 2012). Values in parentheses are the uncertainties in the last digits. Numbers without uncertainties are exact as defined.

Symbol	Meaning	Best Value	Approximate Value
m_e	Electron mass	$9.10938291(40) \times 10^{-31}\text{kg}$	$9.11 \times 10^{-31}\text{kg}$
m_p	Proton mass	$1.672621777(74) \times 10^{-27}\text{kg}$	$1.6726 \times 10^{-27}\text{kg}$
m_n	Neutron mass	$1.674927351(74) \times 10^{-27}\text{kg}$	$1.6749 \times 10^{-27}\text{kg}$
u	Atomic mass unit	$1.660538921(73) \times 10^{-27}\text{kg}$	$1.6605 \times 10^{-27}\text{kg}$

Submicroscopic Masses^[footnote]

Stated values are according to the National Institute of Standards and Technology Reference on Constants, Units, and Uncertainty, www.physics.nist.gov/cuu (accessed May 18, 2012). Values in parentheses are the uncertainties in the last digits. Numbers without uncertainties are exact as defined.

Sun	mass	$1.99 \times 10^{30}\text{kg}$
	average radius	$6.96 \times 10^8\text{m}$
	Earth-sun distance (average)	$1.496 \times 10^{11}\text{m}$
Earth	mass	$5.9736 \times 10^{24}\text{kg}$
	average radius	$6.376 \times 10^6\text{m}$
	orbital period	$3.16 \times 10^7\text{s}$

Epsilon	Ε	ε	Lambda	Λ	λ	Rho	Ρ	ρ	Psi	Ψ	ψ
Zeta	Ζ	ζ	Mu	Μ	μ	Sigma	Σ	σ	Omega	Ω	ω

The Greek Alphabet

	Entity	Abbreviation	Name
Fundamental units	Length	m	meter
	Mass	kg	kilogram
	Time	s	second
	Current	A	ampere
Supplementary unit	Angle	rad	radian
Derived units	Force	$N = kg \cdot m/s^2$	newton
	Energy	$J = kg \cdot m^2/s^2$	joule
	Power	$W = J/s$	watt
	Pressure	$Pa = N/m^2$	pascal
	Frequency	$Hz = 1/s$	hertz
	Electronic potential	$V = J/C$	volt
	Capacitance	$F = C/V$	farad
	Charge	$C = s \cdot A$	coulomb
	Resistance	$\Omega = V/A$	ohm

	Entity	Abbreviation	Name
	Magnetic field	$T = N/(A \cdot m)$	tesla
	Nuclear decay rate	$Bq = 1/s$	becquerel

SI Units

Length	1 inch (in.) = 2.54 cm (exactly)
	1 foot (ft) = 0.3048 m
	1 mile (mi) = 1.609 km
Force	1 pound (lb) = 4.448 N
Energy	1 British thermal unit (Btu) = 1.055×10^3 J
Power	1 horsepower (hp) = 746 W
Pressure	$1 \text{ lb/in}^2 = 6.895 \times 10^3$ Pa

Selected British Units

Length	1 light year (ly) = 9.46×10^{15} m
	1 astronomical unit (au) = 1.50×10^{11} m
	1 nautical mile = 1.852 km
	1 angstrom(\AA) = 10^{-10} m
Area	1 acre (ac) = 4.05×10^3 m ²
	1 square foot (ft ²) = 9.29×10^{-2} m ²
	1 barn (<i>b</i>) = 10^{-28} m ²
Volume	1 liter (<i>L</i>) = 10^{-3} m ³

	1 U.S. gallon (gal) = $3.785 \times 10^{-3} \text{ m}^3$
Mass	1 solar mass = $1.99 \times 10^{30} \text{ kg}$
	1 metric ton = 10^3 kg
	1 atomic mass unit (u) = $1.6605 \times 10^{-27} \text{ kg}$
Time	1 year (y) = $3.16 \times 10^7 \text{ s}$
	1 day (d) = 86,400 s
Speed	1 mile per hour (mph) = 1.609 km/h
	1 nautical mile per hour (naut) = 1.852 km/h
Angle	1 degree ($^\circ$) = $1.745 \times 10^{-2} \text{ rad}$
	1 minute of arc ($'$) = 1/60 degree
	1 second of arc ($''$) = 1/60 minute of arc
	1 grad = $1.571 \times 10^{-2} \text{ rad}$
Energy	1 kiloton TNT (kT) = $4.2 \times 10^{12} \text{ J}$
	1 kilowatt hour ($\text{kW} \cdot \text{h}$) = $3.60 \times 10^6 \text{ J}$
	1 food calorie (kcal) = 4186 J
	1 calorie (cal) = 4.186 J
	1 electron volt (eV) = $1.60 \times 10^{-19} \text{ J}$
Pressure	1 atmosphere (atm) = $1.013 \times 10^5 \text{ Pa}$
	1 millimeter of mercury (mm Hg) = 133.3 Pa
	1 torricelli (torr) = 1 mm Hg = 133.3 Pa
Nuclear decay rate	1 curie (Ci) = $3.70 \times 10^{10} \text{ Bq}$

Other Units

Circumference of a circle with radius r or diameter d	$C = 2\pi r = \pi d$
Area of a circle with radius r or diameter d	$A = \pi r^2 = \pi d^2/4$
Area of a sphere with radius r	$A = 4\pi r^2$

Volume of a sphere with radius r

$$V = \frac{4}{3} \pi r^3$$

Useful Formulae

Glossary of Key Symbols and Notation

In this glossary, key symbols and notation are briefly defined.

Symbol	Definition
any symbol	average (indicated by a bar over a symbol— e.g., \bar{v} is average velocity)
$^{\circ}\text{C}$	Celsius degree
$^{\circ}\text{F}$	Fahrenheit degree
//	parallel
\perp	perpendicular
\propto	proportional to
\pm	plus or minus

Symbol	Definition
0	zero as a subscript denotes an initial value
α	alpha rays
α	angular acceleration
α	temperature coefficient(s) of resistivity
β	beta rays
β	sound level
β	volume coefficient of expansion
β^{-}	electron emitted in nuclear beta decay
β^{+}	positron decay
γ	gamma rays

Symbol	Definition
γ	surface tension
$\gamma = 1/\sqrt{1 - v^2/c^2}$	a constant used in relativity
Δ	change in whatever quantity follows
δ	uncertainty in whatever quantity follows
ΔE	change in energy between the initial and final orbits of an electron in an atom
ΔE	uncertainty in energy
Δm	difference in mass between initial and final products
ΔN	number of decays that occur
Δp	change in momentum

Symbol	Definition
Δp	uncertainty in momentum
ΔPE_g	change in gravitational potential energy
$\Delta \theta$	rotation angle
Δs	distance traveled along a circular path
Δt	uncertainty in time
Δt_0	proper time as measured by an observer at rest relative to the process
ΔV	potential difference
Δx	uncertainty in position
ϵ_0	permittivity of free space
η	viscosity

Symbol	Definition
θ	angle between the force vector and the displacement vector
θ	angle between two lines
θ	contact angle
θ	direction of the resultant
θ_b	Brewster's angle
θ_c	critical angle
κ	dielectric constant
λ	decay constant of a nuclide
λ	wavelength
λ_n	wavelength in a medium

Symbol	Definition
μ_0	permeability of free space
μ_k	coefficient of kinetic friction
μ_s	coefficient of static friction
ν_e	electron neutrino
π^+	positive pion
π^-	negative pion
π^0	neutral pion
ρ	density
ρ_c	critical density, the density needed to just halt universal expansion
ρ_{fl}	fluid density

Symbol	Definition
ρ_{obj}	average density of an object
ρ/ρ_{w}	specific gravity
τ	characteristic time constant for a resistance and inductance (RL) or resistance and capacitance (RC) circuit
τ	characteristic time for a resistor and capacitor (RC) circuit
τ	torque
Υ	upsilon meson
Φ	magnetic flux
ϕ	phase angle
Ω	ohm (unit)
ω	angular velocity

Symbol	Definition
A	ampere (current unit)
A	area
A	cross-sectional area
A	total number of nucleons
a	acceleration
a_B	Bohr radius
a_c	centripetal acceleration
a_t	tangential acceleration
AC	alternating current
AM	amplitude modulation

Symbol	Definition
atm	atmosphere
B	baryon number
B	blue quark color
B	antiblack (yellow) antiquark color
b	quark flavor bottom or beauty
B	bulk modulus
B	magnetic field strength
B_{int}	electron's intrinsic magnetic field
B_{orb}	orbital magnetic field
BE	binding energy of a nucleus—it is the energy required to completely disassemble it into separate protons and neutrons

Symbol	Definition
BE/A	binding energy per nucleon
Bq	becquerel—one decay per second
C	capacitance (amount of charge stored per volt)
C	coulomb (a fundamental SI unit of charge)
C_p	total capacitance in parallel
C_s	total capacitance in series
CG	center of gravity
CM	center of mass
c	quark flavor charm
c	specific heat

Symbol	Definition
c	speed of light
Cal	kilocalorie
cal	calorie
COP_{hp}	heat pump's coefficient of performance
COP_{ref}	coefficient of performance for refrigerators and air conditioners
$\cos \theta$	cosine
$\cot \theta$	cotangent
$\csc \theta$	cosecant
D	diffusion constant
d	displacement

Symbol	Definition
d	quark flavor down
dB	decibel
d_i	distance of an image from the center of a lens
d_o	distance of an object from the center of a lens
DC	direct current
E	electric field strength
ε	emf (voltage) or Hall electromotive force
emf	electromotive force
E	energy of a single photon
E	nuclear reaction energy

Symbol	Definition
E	relativistic total energy
E	total energy
E_0	ground state energy for hydrogen
E_0	rest energy
EC	electron capture
E_{cap}	energy stored in a capacitor
Eff	efficiency—the useful work output divided by the energy input
Eff _C	Carnot efficiency
E_{in}	energy consumed (food digested in humans)
E_{ind}	energy stored in an inductor

Symbol	Definition
E_{out}	energy output
e	emissivity of an object
e^+	antielectron or positron
eV	electron volt
F	farad (unit of capacitance, a coulomb per volt)
F	focal point of a lens
F	force
F	magnitude of a force
F	restoring force
F_{B}	buoyant force

Symbol	Definition
F_c	centripetal force
F_i	force input
\mathbf{F}_{net}	net force
F_o	force output
FM	frequency modulation
f	focal length
f	frequency
f_0	resonant frequency of a resistance, inductance, and capacitance (RLC) series circuit
f_0	threshold frequency for a particular material (photoelectric effect)

Symbol	Definition
f_1	fundamental
f_2	first overtone
f_3	second overtone
f_B	beat frequency
f_k	magnitude of kinetic friction
f_s	magnitude of static friction
G	gravitational constant
G	green quark color
\bar{G}	antigreen (magenta) antiquark color

Symbol	Definition
g	acceleration due to gravity
g	gluons (carrier particles for strong nuclear force)
h	change in vertical position
h	height above some reference point
h	maximum height of a projectile
h	Planck's constant
hf	photon energy
h_i	height of the image
h_o	height of the object
I	electric current

Symbol	Definition
I	intensity
I	intensity of a transmitted wave
I	moment of inertia (also called rotational inertia)
I_0	intensity of a polarized wave before passing through a filter
I_{ave}	average intensity for a continuous sinusoidal electromagnetic wave
I_{rms}	average current
J	joule
J/Ψ	Joules/psi meson
K	kelvin
k	Boltzmann constant

Symbol	Definition
k	force constant of a spring
K_{α}	x rays created when an electron falls into an $n = 1$ shell vacancy from the $n = 3$ shell
K_{β}	x rays created when an electron falls into an $n = 2$ shell vacancy from the $n = 3$ shell
kcal	kilocalorie
KE	translational kinetic energy
KE + PE	mechanical energy
KE_e	kinetic energy of an ejected electron
KE_{rel}	relativistic kinetic energy
KE_{rot}	rotational kinetic energy
KE	thermal energy

Symbol	Definition
kg	kilogram (a fundamental SI unit of mass)
L	angular momentum
L	liter
L	magnitude of angular momentum
L	self-inductance
ℓ	angular momentum quantum number
L_{α}	x rays created when an electron falls into an $n = 2$ shell from the $n = 3$ shell
L_e	electron total family number
L_{μ}	muon family total number
L_{τ}	tau family total number

Symbol	Definition
L_f	heat of fusion
L_f and L_v	latent heat coefficients
L_{orb}	orbital angular momentum
L_s	heat of sublimation
L_v	heat of vaporization
L_z	z - component of the angular momentum
M	angular magnification
M	mutual inductance
m	indicates metastable state
m	magnification

Symbol	Definition
m	mass
m	mass of an object as measured by a person at rest relative to the object
m	meter (a fundamental SI unit of length)
m	order of interference
m	overall magnification (product of the individual magnifications)
$m(^AX)$	atomic mass of a nuclide
MA	mechanical advantage
m_e	magnification of the eyepiece
m_e	mass of the electron
m_ℓ	angular momentum projection quantum number

Symbol	Definition
m_n	mass of a neutron
m_o	magnification of the objective lens
mol	mole
m_p	mass of a proton
m_s	spin projection quantum number
N	magnitude of the normal force
N	newton
N	normal force
N	number of neutrons
n	index of refraction

Symbol	Definition
n	number of free charges per unit volume
N_A	Avogadro's number
N_r	Reynolds number
$N \cdot m$	newton-meter (work-energy unit)
$N \cdot m$	newtons times meters (SI unit of torque)
OE	other energy
P	power
P	power of a lens
P	pressure
p	momentum

Symbol	Definition
p	momentum magnitude
p	relativistic momentum
\mathbf{p}_{tot}	total momentum
\mathbf{p}'_{tot}	total momentum some time later
P_{abs}	absolute pressure
P_{atm}	atmospheric pressure
P_{atm}	standard atmospheric pressure
PE	potential energy
PE _{el}	elastic potential energy
PE _{elec}	electric potential energy

Symbol	Definition
PE_s	potential energy of a spring
P_g	gauge pressure
P_{in}	power consumption or input
P_{out}	useful power output going into useful work or a desired, form of energy
Q	latent heat
Q	net heat transferred into a system
Q	flow rate—volume per unit time flowing past a point
$+Q$	positive charge
$-Q$	negative charge

Symbol	Definition
q	electron charge
q_p	charge of a proton
q	test charge
QF	quality factor
R	activity, the rate of decay
R	radius of curvature of a spherical mirror
R	red quark color
R	antired (cyan) quark color
R	resistance
R	resultant or total displacement

Symbol	Definition
R	Rydberg constant
R	universal gas constant
r	distance from pivot point to the point where a force is applied
r	internal resistance
r_{\perp}	perpendicular lever arm
r	radius of a nucleus
r	radius of curvature
r	resistivity
r or rad	radiation dose unit
rem	roentgen equivalent man

Symbol	Definition
rad	radian
RBE	relative biological effectiveness
RC	resistor and capacitor circuit
rms	root mean square
r_n	radius of the n th H-atom orbit
R_p	total resistance of a parallel connection
R_s	total resistance of a series connection
R_s	Schwarzschild radius
S	entropy
S	intrinsic spin (intrinsic angular momentum)

Symbol	Definition
S	magnitude of the intrinsic (internal) spin angular momentum
S	shear modulus
S	strangeness quantum number
s	quark flavor strange
s	second (fundamental SI unit of time)
s	spin quantum number
s	total displacement
$\sec \theta$	secant
$\sin \theta$	sine
s_z	z-component of spin angular momentum

Symbol	Definition
T	period—time to complete one oscillation
T	temperature
T_c	critical temperature—temperature below which a material becomes a superconductor
T	tension
T	tesla (magnetic field strength B)
t	quark flavor top or truth
t	time
$t_{1/2}$	half-life—the time in which half of the original nuclei decay
$\tan \theta$	tangent
U	internal energy

Symbol	Definition
u	quark flavor up
u	unified atomic mass unit
u	velocity of an object relative to an observer
u'	velocity relative to another observer
V	electric potential
V	terminal voltage
V	volt (unit)
V	volume
v	relative velocity between two observers
v	speed of light in a material

Symbol	Definition
\mathbf{v}	velocity
\mathbf{v}	average fluid velocity
$V_B - V_A$	change in potential
\mathbf{v}_d	drift velocity
V_p	transformer input voltage
V_{rms}	rms voltage
V_s	transformer output voltage
\mathbf{v}_{tot}	total velocity
v_w	propagation speed of sound or other wave
\mathbf{v}_w	wave velocity

Symbol	Definition
W	work
W	net work done by a system
W	watt
w	weight
w_{fl}	weight of the fluid displaced by an object
W_{c}	total work done by all conservative forces
W_{nc}	total work done by all nonconservative forces
W_{out}	useful work output
X	amplitude
X	symbol for an element

Symbol	Definition
${}_A^Z X_N$	notation for a particular nuclide
x	deformation or displacement from equilibrium
x	displacement of a spring from its undeformed position
x	horizontal axis
X_C	capacitive reactance
X_L	inductive reactance
x_{rms}	root mean square diffusion distance
y	vertical axis
Y	elastic modulus or Young's modulus
Z	atomic number (number of protons in a nucleus)

Symbol	Definition
Z	impedance